PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C07H 15/12, C07K 15/14
G01N 33/48, 33/566, C12Q 1/68
C12P 19/34, C12N 5/00, 15/00
C12N 1/20

A1

(11) International Publication Number: **WO 92/12987**

(43) International Publication Date: 6 August 1992 (06.08.92)

(21) International Application Number: PCT/US92/00282

(22) International Filing Date: 10 January 1992 (10.01.92)

(30) Priority data:
639,453 10 January 1991 (10.01.91) US

(60) Parent Application or Grant
(63) Related by Continuation
US 639,453 (CIP)
Filed on 10 January 1991 (10.01.91)

(71) Applicant *(for all designated States except US):* UNITED STATES OF AMERICA as represented by THE SECRETARY DEPARTMENT OF HEALTH AND HUMAN SERVICES [US/US]; National Institute of Health, Office of Technology Transfer, Patent Branch, Bethesda, MD 20892 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only) :* OWENS, Ida, S. [US/US]; 14713 Lake Terrace, Rockville, MD 20853 (US). RITTER, Joseph, K. [US/US]; 10222 Colesville Road, Silver Spring, MD 20901 (US).

(74) Agents: SCOTT, Watson, T. et al.; Cushman, Darby & Cushman, Eleventh Floor, 1615 L Street, N.W., Washington, DC 20036 (US).

(81) Designated States: AT (European patent), AU, BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, LU (European patent), MC (European patent), NL (European patent), SE (European patent), US.

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: THE GENETIC LOCUS UGT1 AND A MUTATION THEREIN

(57) Abstract

The subject invention relates to the isolated gene locus UGT1, a family of UDP-glucuronosyl transferase isozymes encoded by this locus, and to the uses of both the locus as well as the isozymes themselves. In particular, the locus encodes at least six transferase isoforms two of which metabolize bilirubin. The six isoforms share four common exons. The present inventors have discovered that in a patient having Crigler-Najjar Syndrome (CN) Type I, a mutation is present in the second common exon.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**FI**	Finland	**ML**	Mali
AU	Australia	**FR**	France	**MN**	Mongolia
BB	Barbados	**GA**	Gabon	**MR**	Mauritania
BE	Belgium	**GB**	United Kingdom	**MW**	Malawi
BF	Burkina Faso	**GN**	Guinea	**NL**	Netherlands
BG	Bulgaria	**GR**	Greece	**NO**	Norway
BJ	Benin	**HU**	Hungary	**PL**	Poland
BR	Brazil	**IE**	Ireland	**RO**	Romania
CA	Canada	**IT**	Italy	**RU**	Russian Federation
CF	Central African Republic	**JP**	Japan	**SD**	Sudan
CG	Congo	**KP**	Democratic People's Republic of Korea	**SE**	Sweden
CH	Switzerland			**SN**	Senegal
CI	Côte d'Ivoire	**KR**	Republic of Korea	**SU**	Soviet Union
CM	Cameroon	**LI**	Liechtenstein	**TD**	Chad
CS	Czechoslovakia	**LK**	Sri Lanka	**TG**	Togo
DE	Germany	**LU**	Luxembourg	**US**	United States of America
DK	Denmark	**MC**	Monaco		
ES	Spain	**MG**	Madagascar		

THE GENETIC LOCUS UGT1 AND A MUTATION THEREIN

BACKGROUND OF THE INVENTION

Technical Field

The subject invention relates to the isolated gene locus UGT1, a family of UDP-glucuronosyl transferase isozymes encoded by this locus, and to the uses of both the locus as well as the isozymes themselves. In particular, the locus encodes at least six transferase isoforms two of which metabolize bilirubin. The six isoforms share four common exons. The present inventors have discovered that in a patient having Crigler-Najjar Syndrome (CN) Type I, a mutation is present in the second common exon.

Background Information

Bilirubin, an extremely lipophilic-like and toxic metabolite formed by the ongoing turnover of hemoproteins (most significantly hemoglobin) is normally metabolized to excretable derivatives by conjugation of one or both of the carboxyl groups of its propionyl side chains (C-8 or C-12) with glucuronic acid (Ostrow et al., Biochem. J. 120:311-327 (1970)). Both the formation of the C-8 or C-12 monoglucuronide, or the C-8, C-12 diglucuronide, is catalyzed by hepatic bilirubin UDP-glucurono-syltransferase (Ritter, et al., J. Biol. Chem. 266:1043-1047 (1991)). Inadequate bilirubin glucuronidation is commonly observed in neonates (Roy Chowdhury, et al. in The Metabolic Basis of Inherited Disease 1367-1408 (Scriver et al., eds.), McGraw-Hill, New York (1989)), breast-fed infants

(Foliot, et al., Pediatr. Res. 10:594-598 (1976)) and as a consequence of liver diseases (Bloomer, et al., Am. J. Dig. Dis. 19:9-14 (1974)).

Unconjugated bilirubin in cholestatic liver diseases is associated with the accumulation of serum bilirubin and the appearance of the clinical features of jaundice. Coincident with the continuous and marked elevations (>20 mg%) in serum bilirubin concentrations seen in serious diseased states is the risk of kernicterus, the degenerative changes associated with bilirubin deposition in the brain (Roy Chowdhury, et al. 1367-1408 (1989)) in The Metabolic Basis of Inherited Disease, supra (1989). Milder elevations in the serum level of this heme-derivative are seen in approximately half of all neonates causing clinical jaundice which is due to delays in the developmental onset of bilirubin transferase activity normally occurring 1 to 5 days after birth.

Although the short-term maintenance of jaundiced neonates is a major concern of pediatric clinicians, children with inborn errors in bilirubin glucuronidation (totally lacking this function) present an even greater challenge.

Heritable disorders of bilirubin metabolism result in hyperbilirubinemia. These disorders include those resulting in predominantly unconjugated hyperbilirubinemia: Gilbert's syndrome and Crigler-Najjar syndrome, Types I and II. Deficient activity of bilirubin glucurono-syltransferase has been described in each of the three unconjugated hyperbilirubinemias. Bilirubin glucuronosyltransferase is absent in Crigler-Najjar

syndrome, Type I and is partially deficient in Crigler-Najjar syndrome, Type II, and Gilbert's syndrome.

Approximately 6% of the population suffer from Gilbert's syndrome. These patients are typically young adults who present with mild, predominantly unconjugated hyperbilirubinemia, having serum bilirubin levels which are approximately 5-fold elevated over those of individuals not suffering from the syndrome. These patients are always borderline jaundiced (jaundice is a reflection of elevated bilirubin), but are asymptomatic in the absence of other medical problems, such as influenza or biliary artresion. Previously, there has been no way to establish Gilbert's syndrome absent liver biopsy to measure bilirubin levels (which may alternatively be elevated due, for example, to cirrhosis of the liver or hepatitis). Thus, these patients usually suffer high morbidity and incur great expense, having seen many physicians and having undergone many diagnostic tests. Thus, a simple assay specific for bilirubin UDP-glucuronosyltransferase is needed to efficiently and reliably diagnose Gilbert's syndrome. Administration of phenobarbital or other microsomal enzyme inducers is known to reduce the hyperbilirubinemia in Gilbert's syndrome.

In contrast to patients suffering from Gilbert's syndrome, most patients suffering from Crigler-Najjar syndrome, Type I, die between the ages of 3 and 20. This syndrome is characterized by chronic nonhemolytic unconjugated hyperbilirubinemia in which hepatic bilirubin UDP-glucurono-

syltransferase activity is absent (Crigler et al., Pediatrics 10:169-180 (1952)). Crigler Najjar syndrome (CN) Type I, is inherited in an autosomal recessive pattern (Childs et al., Pediatrics, 23:903 (1959); Szabo et al., Acta Paediatr. Hung., 4:153 (1963); and Arias et al., Am. J. Med. 47:395-409 (1969)). The cause of death is the deposit of unconjugated, lipid-soluble bilirubin in the gray matter of the Central Nervous System, leading to seizures, kernicteric damage, and ultimately death. The only effective long-term treatment that has been used successfully in patients with Crigler-Najjar syndrome, Type I, is liver transplantation, although this procedure is not without risk in these individuals. The fatalities associated with this condition confirm the need to detoxify bilirubin. Characterization of the bilirubin transferase gene is, therefore, critical to making progress in the development of therapies for the Type I disease.

Crigler-Najjar syndrome, Type II, is phenotypically similar to Crigler-Najjar syndrome, Type I, except that it is almost always clinically benign and the serum bilirubin concentration is usually below 20 mg/dl. Administration of phenobarbital or other microsomal enzyme inducers is known to reduce the hyperbilirubinemia in Crigler-Najjar syndrome, Type II.

Obviously, in view of the above discussion, there is a need to understand the locus encoding the UDP-glucuronosyltransferase isozymes, a need fulfilled by the present invention.

All U.S. patents and publications referred to herein are hereby incorporated by reference.

SUMMARY OF THE INVENTION

The present invention relates to an isolated gene locus, referred to as UGT1, comprising the nucleotide sequence shown in Figure 1. Within this locus are 6 transcriptional units or DNA segments referred to as UGT1A, UGT1BP, UGT1C, UGT1D, UGT1E, and UGT1F. Each segment encodes a unique transferase isoform.

The present invention also encompasses recombinant DNA molecules comprising a vector and exon 1 of each of the 6 DNA segments.

Another aspect of the invention relates to a DNA segment that codes for a polypeptide having an amino acid sequence corresponding to human bilirubin UDP-glucuronosyltransferase type I or to human bilirubin UDP-glucuronosyltransferase type II.

The present invention also includes the human liver cDNA segment HUG-Br1, ATCC accession under 68510, and the human liver cDNA segment HUG-Br2, ATCC accession under 68509.

Additionally, the present invention encompasses recombinant DNA molecules comprising all of the DNA segments referred to above, as well as procaryotic or eukaryotic cells transformed or transfected with these recombinant DNA molecules.

The present invention also includes uses of the above-described clones or DNA segments. For example, the invention relates to a diagnostic probe of at least 18 bases for the diagnosis of a syndrome

selected from the group consisting of Gilbert's syndrome, Crigler-Najjar syndrome Type I, and Crigler-Najjar syndrome Type II comprising a specific sequence of a cDNA encoding a mammalian bilirubin UDP-glucuronosyltransferase enzyme, mutant form, or variant thereof, said diagnostic probe identifying the corresponding mammalian genomic DNA.

The invention also encompasses, in particular, a diagnostic probe for the detection of Crigler-Najjar, Type I syndrome, wherein said probe comprises a DNA sequence corresponding to the 13-bp nucleotide deletion present in exon 2 of the UGT1 complex.

The invention also relates to a diagnostic probe of at least 18 bases for the diagnosis of a syndrome selected from the group consisting of Gilbert's syndrome, Crigler-Najjar syndrome Type I and Crigler-Najjar syndrome Type II comprising a DNA sequence having sufficient homology to a cDNA encoding a mammalian bilirubin UDP-glucurono-syltransferase form, mutant form, or variant thereof to identify the mammalian genomic DNA corresponding to said cDNA.

Furthermore, the invention also includes a PCR primer pair designed to amplify a specific portion of mammalian genomic DNA, wherein each member of said primer pair is from 17 to 20 bases, and further wherein each member of said primer pair comprises a specific sequence of a cDNA encoding a mammalian bilirubin UDP-glucuronosyltransferase form, mutant form, or variant thereof.

Additionally, the invention includes a diagnostic assay for a syndrome selected from the

group consisting of Gilbert's disease, Crigler--Najjar syndrome Type I and Crigler-Najjar syndrome Type II comprising the steps of:

(a) amplifying a sequence of mammalian genomic DNA using, for example, the PCR primer pair described above to produce multiple DNA copies of said mammalian genomic DNA, and

(b) exposing said amplified DNA according to (a) one of the diagnostic probes described above under hybridization conditions to identify a specific mammalian genomic sequence characteristic of Gilbert's disease, Crigler-Najjar syndrome Type I or Crigler-Najjar syndrome Type II.

Alternatively, one could amplify normal and defective genomic DNA using PCR with one of a pair of normal (i.e., not altered to match the defect) primers located over the defect. PCR product would be expected from template genomic DNA from a normal individual and from both heterozygous parents, but not from genomic DNA from the homozygous defective individual. The failure to generate product with the homozygous defective genome is diagnostic.

Thus, in view of the above, the invention includes a diagnostic assay for the detection of Crigler-Najjar, Type I syndrome in a patient comprising the steps of:

a) altering one of a pair of normal PCR primers such that said altered primer contains the deletion mutation present in exon 2 of UGT1F of the genome of Crigler-Najjar, Type I patients;

b) adding said primers to a DNA sample of a patient suspecting of having said syndrome; and

c) determining whether hybridization occurs such that a PCR product results, lack of said product indicating presence of Crigler-Najjar, Type I syndrome.

It should be noted that the same type of assay could potentially be utilized with respect to the diagnosis of any type of syndrome or disease characterized by a deletion mutation in a gene.

Additionally, the present invention includes the amino acid sequences as shown in Figure 9. One sequence corresponds to human bilirubin UDP-glucuronosyltransferase, Type I (HUG-Br1). The other sequence corresponds to human bilirubin UDP-glucuronosyltransferase, Type II (HUG-Br2).

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1. Nucleotide sequence of the exons, intron-exon injunctions, and 5'- and 3'-flanking regions of the human UGT1 gene complex. The exonic sequences (in capitals) were determined by using a nested sequencing strategy with double-stranded plasmid DNA as template as described in Ritter et al., supra (1990) and by comparison to the HUG-Br1, HUG-Br2, and HLUG P1 cDNA sequences. The TATA boxes are underlined, and the polyadenylation consensus signal sequence is in bold. The predicted amino acid translations (in single letter code) of the open reading frame contained in each exon 1 of UGT1A through UGT1F, and UGT1 common exons, 2-5, are shown below the nucleotide data.

Figure 2. A, represents a schematic diagram of the UGT1 gene locus showing the exon/intron arrangement. UGT1 is a complex of nested transcription units featuring multiple first exons of UGT1A through UGT1F and their associated promoters (right-angled arrows), and four common exons (2-5). The solid black lines indicate intervening DNA. The complex is not drawn to scale and covers at least 85-kb region of chromosome 2 (Harding et al., Ann. Hum. Genet. 54:17-21 (1990)). First exons are spaced an average of 7-kb from each other and the four common exons are clustered in a 6-kb region at the 3' end of the complex.

B, represents the predicted primary transcripts produced from the UGT1 gene locus. Wavy black lines indicate segments of RNA excised from the transcript during splicing and polyadenylation. The RNAs encode a family of UDP-glucurono-syltransferases with unique amino termini (286-289 residues) and identical carboxyl termini (246 amino acids). The products of UGT1A and UGT1D are the bilirubin UDP-glucuronosyltransferases encoded by HUG-Br1 and HUG Br2, respectively (Ritter et al., J. Biol. Chem. 266:1043 (1991)).

Figure 3. Detailed illustration of the UGT1 gene locus. A, Three overlapping clones [44(39.8 kb), 37-1 (43.1 kb), and 42-1 (36 kb)] were isolated from a leukocyte DNA cosmid library constructed in Sp cos2 (Heilig et al., Nucleic Acids Res., 15:9129 (1987)) and transformed into E. coli 490A cells. Approximately 2 x 10^5 colonies were analyzed by the Southern blot technique via hybridization to the unique region of either HUGBr1

(1-470-bp fragment) or HUG-Br2 (1-450-bp fragment) (Ritter et al., J. Biol. Chem., 266:1043 (1991)). The probes were labelled by random primed synthesis in the presence of [α^{32}P]dCTP. Hybridization conditions used were previously described (Ritter et al., J. Biol. Chem., 266:1043 (1991)). Clones 44 and 37-1 hybridized to the HUGBr2-specific probe and not to the HUG-Br1- or to the common end-one. Clone 42-1 hybridized to the HUGBr1- and the 3' common end-[1836 to 2167 bp of the cDNAs] probes (Ritter et al., J. Biol. Chem. 266:1043 (1991)). The sites for 6 different restriction endonucleases and their relative positions among the three clones were determined as shown in Figure 3B. Fourteen contiguous subclones in pBluescriptSK+ were made from the three overlapping cosmid clones. B, Positive subclones which hybridized to HUG-Br1-, HUG-Br2-, or to the common end-probes were mapped in greater detail and sequenced as previously described (Ritter et al. supra, (1991)). (B,C) Four common exons (solid boxes) encoding the common region of the cDNAs and six exon 1s of UGT1A to UGT1F (patterned boxes specifying uniqueness) were identified and are shown according to their distribution in the cosmid clones. Introns are shown as lines. Endonucleases used were: PstI (P), EcoRI (E), XhoI (Xh), XbaI (X), ClaI (C), BamHI (B), HindIII (H), and SalI (S). Exon 1 of UGT1BP or UGT1C contains a KpnI or a SphI site, respectively, which was used to establish orientation. E, Cap sites for the UGT1A, UGT1D, and UGT1F mRNAs (outlined C or G) were determined by primer extension. The cap sites (bold C) were predicted

for the UGT1BP, UGT1C, hand UGT1 mRNAs based on 90% sequence identity to exon 1 of UGT1D in the coding and in at least 200 bp of the flanking regions. Each of the six unique exon ls contained a 5' TATA box (→). C, According to the maps for endonucleases and sequence data, the UGT1 gene locus kb. D, The nested arrangement of the 6 TATA boxes immediately upstream of a predicted cap site of the mRNA and the location of a donor splice motif at the 3' exon/intron junction of these exons (data not shown) indicate that six transcriptional initiation events by RNA polymerase II take place in a nested fashion. Differential splicing of the exon 1 in each primary transcript to the common exons is predicted to produce 6 different mRNAs.

Figure 4. A, Autoradiograms of Sanger nucleotide sequencing reactions of normal and CN, Type I patient DNA. Plasmid DNA containing the entire UGT1 exon 2 from a normal (A) and the CN, Type I patient FB (B) was sequenced using the Bluescript primer, KS. The sequence ladders were generated by electrophoresis through a denaturing 6% polyacrylamide gel, after which the gels were transferred to blotting paper, dried, and exposed overnight to X-ray film to generate the autoradiogram shown. A, C, T and G are lanes corresponding to separate reactions with added dideoxy derivatives of ATP, CTP, TTP and GTP, respectively. The boundary between the first intron and the second exon is shown. The shaded vertical line designates the 13-bp segment in normal DNA which is deleted from the DNA of the CN, Type I patient.

B, Predicted effect of the 13-bp deletion on the bilirubin UDP-glucuronosyltransferases encoded by UGT1A and UGT1D. In a normal individual, the bilirubin transferases encoded by UGT1A (HUG-Br1) and UGT1D (HUG-Br2) have different N-termini (288 or 289 amino acids, respectively) and identical C-termini (246 amino acids). In CN, Type I patient FB, these proteins are predicted to be severely truncated; N and C, the amino and carboxyl termini of bilirubin transferase; CS (unknown function) and MAD (membrane anchoring domain) are conserved domains present in all known transferases.

Figure 5. PCR analysis of genomic DNA from family members. Samples of PCR reactions generated from genomic DNA (0.1 μg) of the CN, Type I individual, FB (Lane 3), his parents (Lanes 2 and 4), and an unrelated normal individual (Lane 1) with oligonucleotides PXG3 and J127 as primers were analyzed by electrophoresis through nondenaturing polyacrylmide (6%) for 5 hr. at 50V in 0.5X TBE. The gel was stained with ethidium bromide, irradiated with UV, and photographed. The primers amplify a fragment representing the last 45-bp of the first intron and the first 154-bp of exon 2. Lane 5 is a sample of a control reaction in which DNAs from FB (homozygous affected) and the unrelated normal (homozygous normal) were mixed in approximately equal proportions and subjected to the same PCR analysis. STD is a molecular size standard, 1 Kilobase Ladder (Bethesda Research Laboratories). Molecular sizes (in base pairs) are shown to the right of the figure.

Figure 6. A, Northern blot analysis of human liver mRNA and kidney and skin total RNA samples using UGT1A-, UGT1D-, and UGT1F-specific probes. Triplicate sets of two different liver mRNA (2μg each) samples, and one kidney and one skin total RNA (25 μg each) samples were prepared for analysis according to published procedures. (Ritter et al., J. Biol. Chem., 265:7900 (1990)). Corrections for differences in probe lengths were made in calculations of relative amounts of mRNA. A human B-actin probe was hybridized (Church et al., Proc. Natl Acad. Sci USA 81:1991 (1984)) to each filter to normalize the amount of RNA applied to the gel.

B, Southern blot analysis of endonuclease-digested human genomic DNA using three different exon 1- and the common exon 5-specific probes. Human lymphocyte DNA was digested as indicated, electrophoresed in a 0.8% agarose gel, and blotted on to ZetaBind™ membrane to make a duplicate set. Probes are defined at the top of the panel. One filter was first exposed to a UGT1F-probe, stripped, and rehybridized to a UGT1A-specific one. The second filter was exposed to the exon-5 probe and then to a UGT1D one in a similar manner. The hybridization conditions were as described (Ritter et al., supra, (1990)).

Figure 7. The conjugation of glucuronic acid (donated by UDP-glucuronic acid) to bilirubin-IXα to produce water-soluble IXαC12 Bilirubin-ß-glucuronide.

Figure 8. Three different Bilirubin-ß-glucuronide conjugates: bilirubin-IXαC-8

monoconjugate, bilirubin-IXαC-12 monoconjugate, and bilirubin-IXα diconjugate.

Figure 9. The nucleotide and deduced amino acid sequences of HUG-Br1 and HUG-Br2. Nucleotide and deduced amino acid residues in HUG-Br1 which differ from HUG-Br2 are shown in reverse font. The start and stop codons are designated by open boxes. Putative membrane-insertion signal and membrane-anchoring peptides are indicated by dashed and solid lines, respectively. Predicted asparagine-linked glycosylation sites are denoted by solid triangles and consensus polyadenylation signals by dashed boxes.

Figure 10 (Panel A). Hybridization of HUG-Br1 and HUG-Br2 to liver mRNA isolated from normal human and from a Crigler-Najjar syndrome patient. Duplicate sets of mRNA samples (2 ug each) from two normal (N) and a type I Crigler-Najjar patient (CN) were electrophoresed as described (Ritter, et al. J. Biol. Chem. 265: 7900-7906 (1990)) and transferred onto ZetaBind™ membrane. One filter was hybridized to a HUG-Br1-specific probe and one to a HUG-Br2-specific probe. The amount of mRNA applied to the gel was normalized by hybridization to a 660-bp human cyclophilin probe (bottom of panel A). All mRNA levels were quantitated by densitometric scans of appropriately exposed X-ray film, and the areas under the curves were then evaluated by computer analysis.

Figure 10 (Panel B). Hybridization of HUG-Br1 and HUG-Br2 to liver mRNA isolated from the Erythrocebus patas monkey. Duplicate sets of mRNA (2 ug) isolated from either untreated monkey or one

maintained on 15 mg% phenobarbital (PB) in the drinking water were analyzed as in Figure 6A.

Figure 11. Expression of Bilirubin Transferase Activity in Mouse Microsomes and COS-1 Cells Transfected with either pHUG-Br1 and pHUGBr2: Bilirubin glucuronidation was measured using ^{14}C-bilirubin and UDP-glucuronic acid, and the glucuronides were transesterified to their corresponding methyl esters and subjected to TLC. The TLC plates were exposed to X-ray film (10 days). Panels A and B represent two different experiments. Bilirubin glucuronides formed by mouse microsomes (2.9 mg) in the presence (A, Lane 3) and absence (A, Lane 4) of UDP-glucuronic acid are shown. Sensitivity of the reaction product generated in 6.5 hr. to glucuronidase treatment for 1 hr. at 37°C is shown in A, Lane 2. The methyl esters of bilirubin corresponding to the IXαC8 and the IXαC12 isomeric monoglucuronides (BMEs), the diglucuronide (BDE), and unconjugated bilirubin (Br) are designated. In this system, the rank in mobility is: Br > IXαC8 > IXαC12 > IXαC8,IXαC12. Cells transfected with pHUG-Br2 or pHUG-Br1, which contained the insert in the correct orientation with respect to the promoter element in pSVL, are shown in A (Lane 1) and B (Lane 2), respectively. Cell homogenates were assayed as described for microsomes. Control homogenate from cells transfected with pHUG-Br1 in the reversed orientation is shown in B (Lane 3). The sensitivity of the product generated by HUG-Br1-encoded activity to B-glucuronidase is shown in B, Lane 1. Unconjugated bilirubin is more sensitive to oxygen than its glucuronide or methyl ester derivatives,

and the oxidized bilirubin will remain at the origin in this system. A nonspecific band (NS) is present in each reaction.

Figure 12. A comparison of the amino acid sequences of various UDP-glucuronosyltransferase forms: rat phenol UDP-glucuronosyltransferase (Rat PHENOL), human phenol UDP-glucuronosyltransferase (HLUGP1), human bilirubin UDP-glucuronosyltransferase, type I (HUB-BR1), human bilirubin, type II (HUG-BR2), and rat bilirubin UDP-glucuronosyltransferase (Rat Bil). As shown in the figure, the carboxyl terminus of the various UDP-glucuronosyltransferase forms is conserved as between the human and rat species and the phenol and bilirubin acceptor substrate forms. The human forms, HLUGP1, HUG-Br1 and HUG-Br2, show exact homology between amino acids 283-531, 285-533, and 286-534, respectively. The rat forms, Rat PHENOL and Rat Bil, show exact homology between amino acids 281-529 and 283-531, respectively. Further, there is a 90.6% conserved amino acid sequence between the two species. In the nonconserved region of Rat PHENOL, GLUGP 1, HUG-BR1 and HUG-BR2 and Rat Bil, corresponding to amino acids 1-280, 1-282, 1-284, 1-285, and 1-282, respectively, there is 66% homology between HUG-BR1 and HUG-BR2, 70% homology between HUG-BR1 and Rat Bil, 76% homology between HUG-BR2 and Rat Bil, 84% homology between Rat PHENOL and HLUGP1, and 60% homology between Rat Bil and HLUGP1.

Figures 13A and B. Restriction maps of cosmid clones containing the genes coding for HUG-Br1 and HUG-Br2. Note Figure 13A is continuous with Figure 13B, overlapping at the arrows shown on each.

DETAILED DESCRIPTION OF THE INVENTION

The subject invention relates to the UGT1 locus. The nucleotide sequence of UGT1 is shown in Figure 1. UGT1 is a unique gene complex featuring six nested transferase transcriptional units, which are depicted in Figure 2 (panel A). Each of these units, UGT1A-UGT1F, encodes a unique transferase isoform. The complex contains a tandem array of six promoters (Fig. 2A, arrows) with each positioned adjacent to a first exon (Fig. 2A, designated boxes, ≈880-960-bp each), thereby defining the transcriptional start site for the unit.

The six unique amino acid termini of 286-289 amino acids are encoded by the six different first exons and identical carboxyl termini of 246 amino acids are encoded by the common exons 2-5. It is thought that critical mutations introduced into any of the common exons 2-5 will result in the inactivation of all encoded isoforms. Alternatively, a critical mutation in exon 1 or a specific regulatory region can generate a specific defect in that encoded isoform or responsiveness to drug treatment, respectively. Mutations in sequences necessary for RNA splicing (e.g., upstream of a common exon) can generate a defect in all the transferase isoforms. The evidence from the gene structure and the presence of corresponding mRNAs indicate that typical liver microsomal bilirubin UDP glucuronosyltransferase is necessarily composed of at least two distinct isozymes encoded by the cDNAs HUG-Br1 and HUG-Br2 (Ritter et al., J. Biol. Chem.

266:1043-47 (1991)) which are encoded by the UGT1 locus.

The discovery of the two bilirubin transferase cDNAs of the present invention, HUG-Br1 and HUG-Br2, as described below, provided and established for the first time the nucleotide sequence data which code for any human bilirubin transferase enzyme.

In particular, the present invention relates to the isolation and characterization of these two human liver bilirubin UDP-glucurono-syltransferase cDNAs, referred to as HUGBr1 and HUG-Br2 (Ritter, et al., J. Biol. Chem. 266:1043-1047 (1991)) which, upon expression individually in COS-1 cells, encode isoforms that catalyze the formation of the two bilirubin monoglucuronides and the diglucuronide. Northern analysis of HUG-Br1 and HUG-Br2 mRNA showed that HUG-Br1 is 4-fold more abundant than HUG-Br2, and that, in the Erythrocebus patas monkey, mRNA hybridizing with HUG-Br2, but not HUG-Br1, is inducible by phenobarbital treatment. It was also discovered that mRNAs corresponding to both bilirubin transferase isoforms were present in the explanted liver of a CN, Type I patient. (Ritter et al., J. Biol. Chem. 266:1043-1047 (1991)). The observations that the mRNAs were of the normal size (≈2.6kb) and abundance suggest that neither defective regulation nor gross abnormalities in gene structure is causative and that smaller point mutations are probably involved.

Sequence data showed that the cDNAs contained identical 3' ends (1469-bp in length) to each other and to that of the human phenol transferase cDNA, HLUGP1 (Harding et al., Proc. Natl Aca. Sci. USA 85:8381 (1988)). (The phenol isozyme specifically conjugates a series of hydroxyl-substituted aromatic compounds.) In contrast, HLUGP1, HUG-Br1 and HUG-Br2 have unique 5' ends (879, 882- and 912-bp) which encode the first 286 of 534, 288 of 533, and 289- amino acids of 534, respectively. The identity in the 3' region of the 3 clones suggested that a gene arrangement exists such that primary transcripts containing unique exons are alternatively spliced to one or more common exons to account for the structural similarities. Sato et al. (Biochem. Biophys. Res. Comm. 169:260 (1990)) have shown that a rat bilirubin transferase cDNA contains a 3' end identical to that of the rat phenol transferase cDNA (Iyanagi et al., J. Biol. Chem. 261:15607 (1986)). The carboxyl-termini between rat and human isoforms are 90% identical.

Furthermore, prior to these findings, it was not known that more than one bilirubin metabolizing isoform exists and that one of these is responsive to phenobarbital treatment. Phenobarbital has been the agent for treating and/or diagnosing patients with certain defects with bilirubin glucuronidating activity. The cDNAs provided the necessary probes for isolating, sequencing and describing the genes encoding the isoforms. The purification and nucleotide sequencing of the genes encoding the bilirubin and

phenol transferase isoforms and all flanking regions provide all the necessary information to describe and verify all critical and non-critical mutations which may lead to an impairment or total loss of bilirubin glucuronidating activity in affected individuals. The gene complex arrangement with an independent promoter for each encoded isoform and the sharing of 4 common exons provide a basis for the human phenotypes seen with respect to impaired bilirubin and phenol transferase activities. The sequence data of the normal gene allows one to amplify and sequence all relevant regions of a defective gene to determine alterations and to confirm the effect of the mutation on activity by altering the cDNA before expression in an appropriate host cell. Prior to the present invention, there was no available gene-based method for approaching or developing an analysis of abnormal detoxification of bilirubin in humans.

Primary messenger RNA transcripts are predicted by each unit as shown in Figure 2 (panel B) and are further predicted to contain different 5' termini but identical 3' termini. The model indicates that each transcript undergoes differential splicing of the first exon to exon 2 (common exon) and then is further extended with exons 3, 4, and 5 (common exons) to produce mature RNAs approximately 2.6-kb in length.

Two of the 6 transcriptional units, UGT1A and UGT1D, respectively, represent the transcriptional units which encode mRNAs corresponding to HUG-Br1 and HUG-Br2 bilirubin transferase cDNAs prepared from mRNA isolated from

normal human liver. A third unit, UGT1F, encodes a phenol-specific isoform (Dutton, Glucuronidation of Drugs and Other Compounds, CRC Press, Boca Raton, Fl., pp. 3-78 (1980)).

Exon 1 of UGT1A, UGT1D, and UGT1F is located 5.6-, 49-, and 73-kb, respectively, upstream of the 4 common exons (See Fig. 3B and 3C.) Three additional unique exon 1s of UGT1BP, UGT1C, and UGT1E which are -90% identical to and of the same length as that of the HUG-Br2 exon (UGT1D), but never before uncovered, were identified 19-, 38-, and 54-kb upstream of the 4 common exons (Fig. 3B and 3C). The first exons of UGT1C and UGT1E are predicted to encode the unique regions of proteins which are similar to the HUG-Br2 protein. The first exon of UGT1BP has a frame-shift mutation which creates a premature stop codon and, presumably, a truncated protein. It is, therefore, designated a pseudogene (P). Because of the 90% identity of the exons 1s with that of UGT1D, it is possible that UGT1C and UGT1E encode bilirubin transferase isoforms with properties similar to that encoded by HUG-Br2.

Each of the 6 unique exon 1s of UGT1A to UGT1F has an RNA polymerase II transcriptional promoter element, TATATATATATATAA, TAATTAA, or TATCAAA, upstream of the predicted cap site of the mRNA as detailed in Fig. 3C and 3E. The presence of these elements at position - 37 to -23 bp (UGTIA) or -31 to -25 bp (all other exon 1s) from the cap site is consistent with the typical location of TATA boxes. The presence of individual promoters create the condition whereby each encoded isoform can have

independent regulation and tissue-specific expression. Hence, an independent mutation in a promoter region can specifically affect the expression of that isoform.

Each exon 1 has a 3' exon/intron junction consistent with a donor splice site (data not shown). The common exons 2, 3, and 4 contain an acceptor and donor splice site, common exon 5 contains only an acceptor site. The model predicts the synthesis of at least 6 nested primary transcripts ranging from 15- to 95-kb (Fig. 3D) where each exon 1 is differentially spliced to the 4 common exons to generate 6 different mRNAs.

The exons 1s uncovered in this invention allow one to explain substrate specificity of the isoforms and to explain defects in humans where only bilirubin transferase activity is impaired and not phenol transferase activity. Furthermore, the discovery of other bilirubin transferase-like encoded isoforms (UGT1C and UGT1E) potentially provide other expressible forms with the appropriate drug treatment. Hence, UGT1A (HUG-Br1) and UGT1D (HUG-Br2) are now known isoforms which detoxify bilirubin. Moreover, UGT1C and UGT1E are possibly (based on 90% identity to UGT1D) bilirubin transferase isoforms. The identification of UGT1A, UGT1C, UGT1D, and UGT1E allows one to determine if nucleotide alterations exist in each of the corresponding exon 1s as well as in the 4 common exons in humans with defective bilirubin transferase activity. Furthermore, upon establishing all the types of defects which exist in human subjects, nucleotide oligos can be designed and applied to the

diagnostic screening of this population of individuals.

Preliminary data show that DNA from the explanted liver from a CN, Type I patient has a codon (3 nucleotides) deletion in UGT1A without any alterations in the common exons. This individual conjugated 12% of the bilirubin in the bile, presumably due to a normal activity encoded by the UGT1D mRNA which is expressed at a low level. The RNA studies on samples from this individual have been reported (Ritter et al., J. Biol. Chem. 266:1043-47 (1991)). This individual was reported to have normal phenol transferase activity.

Comparisons of the UGT1 gene complex with the only characterized transferase gene (Mackenzie et al., J. Biol. Chem. 265:11328 (1990)) (rat, testosterone) show that the size and distribution of exons in that simple gene are very similar to those in this complex. For the rat gene, exons 1 and 2 code for the 293-NH_2:-terminal region, and exons 3 to 6 code for 44, 29, 74, and 91 amino acids. For the UGT1 locus, exons 2 to 5 code for 44, 29, 74, and 99 amino acids, respectively. The similarities suggest that other domains are represented in the 3' exons which are not yet understood.

Complementary DNA cloning studies established that the mRNAs encoding UGT1A, UGT1D, and UGT1F are expressed in liver, and induction studies indicate differential responsiveness of promoter elements (Ritter et al., J. Biol. Chem. 266:1043 (1991) and Harding et al., Proc. Natl. Acad. Sci. USA 85:8381 (1988)). Northern blot analysis (Fig. 6A) establishes that the relative

level of mRNA in liver encoding UGT1A, UGT1D, and UGT1F is 5:2:1. (The present inventors have also demonstrated that the mRNA coding for UGT1D, but not UGT1A, is induced in monkey by phenobarbital treatment. (Ritter et al., J. Biol. Chem. 266:1043 (1991)). Preliminary sequence analysis of UGT1D-like cDNAs generated by RT-PCR approach showed that UGTID- and UGTIC-specific poly (A+) RNA account for 88 (44/50) and 12% (6/50), respectively, from a single individual (Harding et al., supra (1988)). Figure 6 shows that UGT1F mRNA is expressed in kidney and skin at levels equal to or higher than that in liver, whereas the mRNAs coding for UGT1A and UGT1D are not detected in kidney and skin. Human liver is the only known site for bilirubin transferase activity (Chowdhury et al., The Metabolic Basis of Inherited Diseases (Scriver et al., eds.) McGraw Hill, N.Y. pp. 1367-1408 (1989)). Thus, the multiple initiation sites and splicing to produce alternative protein products have major consequences for cellular phenotype. The liver selectively processes bilirubin for clearance through the biliary system, while phenol clearance is not restricted and takes place in many different tissues. The presumed similarities in the chemical nature of bilirubin and phenol and/or the biological circumstances leading to the establishment and evolution of a single gene complex to encode detoxifying enzymes are not obvious.

The implication of this gene arrangement is that multiple bilirubin transferase isozymes have evolved to cope with and prevent the toxic effects of the ever-present waste product derived from

hemoprotein turnover. It is known that bilirubin represents the only biological toxin in mammals undergoing constant replenishment: a) from dying red blood cells and, b) from episodic increases due to red blood cell hemolysis. Thus, the toxin most likely represents a powerful selective pressure to evolve an efficient detoxifying system. Duplication events within the UGT1D-like alleles have likely been in response to the need to expand the only mechanism for detoxifying this heme derivative. The syndromes in humans attributed to mild elevations in serum bilirubin are in response to a modest decrease in bilirubin transferase activity. This reciprocal relationship likely signals that the constitutive level of the enzyme is not in excess for metabolizing this endogenous toxin. Furthermore, the inducibility of the mRNA encoiding the UGT1D gene by phenobarbitol expands the glucoronidating capacity and provides greater flexibility to the organism in responding to environmental factors. The gene arrangement is flawed by the obvious prediction that a single critical mutation in a common exon can inactivate the entire subfamily of isozymes. Such a critical mutation is found in the genome of the first patient with Crigler-Najjar, Type I disease to be described.

As supported by Example II, the present inventors have discovered a homozygous 13-bp deletion in common exon 2 of the gene complex, UGT1, in a patient with CN, Type I, a deletion which is related to the claimed invention. The nature of this defect is consistent with inactivation of the gene product. Figure 4B shows the proposed effect

of the frameshift mutation which results in the synthesis of truncated forms of both bilirubin transferase isozymes. Figure 4A shows the site leading to a change in its predicted translation. The FB sequence would, therefore, be very different from that of a normal individual after codon 294. Continuation of translation of the open-reading-frame (not shown in Fig. 4A) shows a TGA stop signal after the 365th codon (illustrated schematically in Fig. 4B); thus, a 72-residue unrelated peptide is present after codon 294.

As deduced by translation of the altered coding region, the bilirubin transferases represented by UGT1A and UGT1D are predicted to be missing the final 240 amino acids of the wild type proteins, a region in the phenol/bilirubin transferase gene which is highly conserved across species (90% identity between rat and human) (Ritter et al. J. Biol. Chem. 266:1043-1047 (1991)). This region contains two smaller domains/motifs that have been recognized as highly conserved among transferase cDNAs encoding members of both major subgroups so far identified, the steroid and bilirubin/phenol subfamilies (Tephly et al., TIPS 11:276-279 (1990)). The membrane-spanning domain at the extreme carboxyl-terminus of all known transferases is missing. The absence of this 17 consecutive hydrophobic residues normally flanked on the amino terminal side by a charged residue (Asp) and the carboxyl terminal side by a highly positively charged halt transfer signal (Lys/Arg rich) should cause the UGT1-encoded proteins to distribute in the cytosol. In addition, the highly

conserved 18-residue motif including the first 6 residues of the membrane spanning domain (present in all transferase isozymes and utilized as a strategy for the cloning of the human (Ritter et al. J. Biol. Chem. 266:1043-1047 (1991)) and rat (Sato et al., Biochem. Biophys. Res. Comm. 169:260-264 (1990)) bilirubin isoforms) is also missing in FB. Its function is currently unknown although it has been speculated that it participates in the binding of the common donor substrate, UDP-glucuronic acid.

In addition to the predicted loss of four codons (294-297), as noted above, the truncated proteins are predicted to contain a foreign 72-residue carboxyl terminus, resulting from the -1 shift in reading frame. A 1-bp deletion with a similar consequence (i.e., the introduction of a premature stop codon) was described at the cDNA levels for phenol and bilirubin transferase mRNAs of the hyperbilirubinemic Gunn rat, considered a model of CN, Type I disease (Schmid et al., J. Clin. Invest., 37:1123-113 (1958)). Iyanagi et al., J. Biol. Chem. 264:21302-21307 (1989)) showed that defective 3-methylcholanthrene-induction of phenol glucuronidating activity was due to a -1 frame shift mutation at codon 414 in the mRNA for the phenol form (Iyanagi, et al. J. Biol. Chem. 261:15607-15614 (1986)) (which is the orthologue of the form encoded by the human UGT1F). Presumably, the gene arrangements are similar between human and rat. This frameshift would, therefore, occur in a common exon of the corresponding locus in the rat. This mutation results in the synthesis of a truncated enzyme (Elwady et al. J. Biol. Chem. 265:10752-10758

(1990)). More recently, Sato et al. showed that the genetic defect in bilirubin UDP-glucurono-syltransferase in the Gunn rat was due to the same mutation (Sato et al., Biochem. Biophys. Res. Comm. 177:1161-64 (1991)). These data together with the recent determination of the UGT1 gene structure provide an understanding of the complex arrangement of the locus and allow the prediction that the other forms encoded within UGT1 are also defective in the CN, Type I patient studied here, including the isozymes encoded by UGT1BP, UGT1C, and UGT1E.

Both Crigler-Najjar Syndrome, Type I, and the hyperbilirubinemia seen in the Gunn rat, are transmitted as autosomal recessive disease entities (Childs et al. Pediatrics. 23:903 (1959); Szabo et al., Acta Paediatr. Hung., 4:153 (1963); and Arias et al., Am. J. Med. 47:395-409 (1969)). In humans, CN, Type I is a rare disorder which was first described in the literature in 1952 (Roy Chowdhury et al., In The Metabolic Basis of Inherited Disease, pp. 1367-1408 (1989)) and is often a consequence of consanguineous parents. As for the case described here, the heterozygosity and consanguinity of the parents are the basis of the homozygosity of FB at the UGT1*FB locus.

Using PCR, the present inventors conclusively demonstrated that both parents are carriers of the same defect seen in FB. Whereas only a shortened band was amplifiable from the DNA of FB, bands of both the deleted size and normal size were amplifiable from that of the parents (Fig. 5). The PCR assay used herein may be utilized as one in a series of diagnostic markers--among others

to emerge from future characterization of the different defects associated with heritable hyperbilirubinemic disorders (e.g. CN, Type II or Gilbert's syndrome).

Although it is known that a 13-basepair deleteion accounts for the UGT1*FB CN, Type I case, the existence of multiple phenotypes associated with CN, Type I (Arias et al., Am. J. Med. 47:395-409 (1969); Robertson et al. J. Inher. Metab. Dis. 14:563-579 (1991)) supports the notion that multiple types of mutations in UGT1 are involved. To date, the present inventors have uncovered at least two other types of alterations in CN, Type I patients: a) a single nucleotide change in common exon 2 which creates an amino acid change at residue 309, a conserved region of the molecule, and b) a codon deletion in exon 1 of UGT1A leading to the deletion of residue 170. Furthermore, Van Es et al. (Van Es, et al., J. Clin. Invest. 85:1199-1205 (1990)) performed a biochemical and immunochemical analysis of four different (unrelated) CN patients with no detectable transferase activity toward bilirubin. Whereas patients A and B had reduced microsomal transferase activity towards 5-hydroxytryptamine and phenol, patients C and D had normal activities toward these substrates. Furthermore, by using two different monoclonal antibodies (WP1 and HEB7) prepared against phenol and bilirubin UDP-glucuronosyltransferase, Van Es et al. showed that two or more immunostained isozymes were absent in patients A and B, whereas in patients D and C, the polypeptide patterns were similar to controls. These results also suggest a genetic heterogeneity

among CN, Type I patients. Some are apparently able to synthesize full length (but presumably inactive) bilirubin transferases, whereas others (such as FB described here) synthesize truncated versions. From the results contained herein, one can predict that microsomes from patient FB (who has not undergone liver transplant surgery and therefore has not yet been tested) will exhibit the phenotype which characterized Patients A and B. It also suggests the possibility that among CN patients mutations are possible which selectively inactivate one or both of the bilirubin isoforms while sparing the remaining forms encoded by the UGT1 locus (e.g. the UGT1F-encoded phenol form). It is expected that analysis of additional CN patients will uncover a host of different mutations. Preliminary finding of a different alteration in exon 2 of a second CN, Type I patient and a codon deletion in exon 1 of UGT1A of a third CN, Type I patient confirm the prediction of the present inventors that many types of genetic alterations account for the CN, Type I phenotype.

The above-described invention has many significant uses. As noted above, the present inventors have determined the sequence of all known exons encoding bilirubin UDP-glucurono-syltransferases within the ≈90 kb span of the normal UGT1 gene complex. Thus, the normal sequence data have allowed the inventors to detect genetic alterations in the genomic DNA from three different Crigler-Najjar, Type I patients. FB has been described in detail in example IV with confirmation that each parent's DNA is heterozygous for the defect. Although two other genetic alterations have

been determined, it has not been confirmed that either the single nucleotide change leading to an amino acid change or a codon deletion leading to an amino acid deletion generates a loss of enzyme activity. with the use of pairs of primers (at least 17 nucleotides in length) in the UGT1 gene complex, one can amplify any target region of the genome from normal and defective DNA by the PCR methodology using both a Perkin-Elmer kit and a thermocycler, for example, to generate sufficient DNA for subcloning. Subclones will be sequenced by the method described above to establish sequence data for comparison to normal by using, for example, the IBI Pustell sequence analysis analysis software package. Each defect will provide a site for a diagnostic probe.

The above-described methodologies can be applied to locate other altered nucleotide sequences in the DNA from other patients with Crigler-Najjar, Type I, Crigler-Najjar, Type II, and Gilbert's syndromes to uncover all mutations or variations in the nucleotide coding sequences or in flanking regions of the UGT1 gene complex and other mammalian bilirubin UDP-glucuronosyltransferase genes.

In order to establish a diagnostic probe, sufficient quantities of the two types of DNA, as well as samples of HUG-Br1 and HUG-Br2 cDNAs, can be analyzed by Southern blot analysis (Jaye et al., Nucleic Acid Research 11:2325-35 (1983)). Each blot may be made in duplicate. An approximately 18-base oligo reflecting the altered sequence in the Crigler-Najjar, Type I, as well as an approximately 18-base oligonucleotide reflecting this sequence in

normal DNA can be made and each can be radiolabelled to at least 2 x 10^8 dpm/ug by T4 polynucleotide kinase and/or terminal transferase to be used as a probe. Preliminary studies can be carried out to establish wash conditions for each normal and altered sequence (see Sambrook et al., In Molecular Cloning-A Laboratory Manual). Following hybridization, high stringency wash conditions (as established for the filters) can be carried out in combination with the test sample of DNA present on each filter. The validity of the diagnostic probe can be established by the selective hybridization (under high stringency wash conditions) of each probe to target DNA containing a perfect match.

The PCR methodology could also be utilized diagnostically in a slightly different manner. One of a pair of normal primers (not altered to match the defect) could be selected to include the defect in the genome. It would be predicted that PCR product could be generated from the genome of the normal individual and that of both heterozygous parents for the defect, but not that homozygous for the defect. The failure to generate product would be diagnostic for a defect at the site of the specified primer.

Similarly, one can analyze genomic DNA from patients with Crigler-Najjar, Type II and Gilbert's syndromes for altered DNA sequence(s) compared to normal. Any altered DNA sequence can serve as a target for a diagnostic probe to be compared to this sequence in normal DNA.

The present invention can also be used with respect to gene therapy. For example, Crigler-

Najjar, Type I disease, a disease which is fatal for most patients between the ages of 3-20, could potentially be treated using the claimed invention. (Liver transplantation has been the only effective long-term treatment.) In particular, the present invention provides two cDNAs which encode two different bilirubin UDP-glucuronosyltransferases which can be used as agents for gene therapy in order to correct the defect in the expression of this activity. Although gene therapy research is in its infancy, progress is being made with animal models. The protocols relevant for the bilirubin transferase gene therapy model necessarily deal with liver. Currently, amphotropic retroviral vectors which have been engineered to be non-disease producing are utilized as agents for the transmission of foreign DNA into cultured primary liver cells prepared from a portion of the liver from an affected animal (individual). After the ex vivo gene replacement, the cells are injected into the spleen of the same animal. From this site, most of the cells migrate to and remain in the liver.

The present invention can be illustrated by the use of the following non-limiting examples:

Example I

Sequencing of CN, Type I Human Samples

The following protocol was utilized in order to sequence the samples of interest: CN, Type I human samples: Blood samples were collected from FB, a 4 year old male CN, Type I patient, his parents, and an unrelated normal female, AR. FB, the offspring of consanguineous parents, had average

serum bilirubin concentrations of 598 μmol/liter which are not affected by phenobarbital therapy. Hemoglobin was 19.6 gm%, and the reticulocyte count was 0.7%. The mother was O negative and the baby was O positive, both direct and indirect Coombs' tests were negative. The results of the following studies were normal: urinary menthol glucuronides following oral administration (Szabo et al., Acta Paediatr. Hung., 4:153 (1963)), transaminases, alkaline phosphatases, 5'-nucleotidase, total bile acids, thyroid function tests, PT, PTT, G6PD screen, pyruvate kinase screen, α-antitrypsin levels, and various bacterial cultures. Unconjugated hyperbilirubinemia persisted despite repeated exchange transfusions and phototherapy. A liver biopsy at 6 weeks of age confirmed normal histology and the absence of bilirubin UDP-glucurono-syltransferase activity. Bilirubin glucuronides were absent in duodenal bile, serum, and urine.

Preparation of lymphocyte genomic DNA:

The blood samples were transferred to 50-ml Falcon tubes and spun at 4000 rpm for 30 min in a HS-4 rotor (Dupont-Sorvall). The pellet was resuspended in 40 ml of prechilled RSB (10 mM Tris-Cl, pH 7.5, 10 mM NaCl, 3 mM MgCl, and 0.2% Triton X-100), recentrifuged, and the step repeated a second time. The pellet was resuspended the final time in 2 ml of SDS solution (10 mM Tris-Cl, pH 7.5, 20 mM NaCl, 5 mM MgCl, 2% SDS). Proteinase K (900 μg/10 ml volume blood) was added, and the tubes were incubated at 37°C for 2 hr and then at room temperature overnight. The samples were then

extracted five times with 1:1 volume of phenol:chloroform and dialyzed at 4°C overnight against 10 mM Tris, pH 7.5, 1 mM EDTA with several changes.

Oligonucleotide primer synthesis: Oligonucleotides for PCR amplification of the four UGT1 common exons [PXG3-PXG4 for exon 2, PXG5 and PXG6 for exon 3, PXG7 and PXG8 for exon 4, PXG9 and PXG10 for the protein coding region of exon 5] were synthesized using a Milligen/Biosearch Cyclone DNA synthesizer, Model 8400 (Milligen/Biosearch, Bedford, MA) and a 200-coupling ß-phosphoamidite kit (Milligen/Biosearch). The oligonucleotides and their sequences are as follows:
PXG3 5'-CTATCTCAAACACGCATGCC-3', PXG4 5'-GGATTAGCGCTCCTGTGAA-3', PXG5 5'-GTCTTTCTTTACGTTCTGCTC-3', PXG6 5'-GACCCTGGTTTGACCTATAC-3', PXG7 5'-CTCAGAGATGTAACTGCTGAC-3', PXG8 5'-CATGAATGCCATGACCAAAG-3', PXG9 5'-GTTCATACCACAGGTGTTCCA-3', PXG10 5'-GGAAATGACTAGGGAATGGTTC-3', J127 5'-TCTGAGACCATTGATCC-3'.
After synthesis, the oligos were cleaved and deprotected with concentrated aqueous ammonia (29%) as suggested by the manufacturer, dried and resuspended in sterile water, re-precipitated with ethanol, and resuspended in 10 mM Tris, pH 8.0, 1 mM EDTA. Oligos were analyzed by electrophoresis through a 12% polyacrylamide gel.
Polymerase chain reaction, subcloning, and

sequencing of PCR products: Primer combinations PXG3 and PXG4, PXG5 and PXG6, PXG7 and PXG8, or PXG9 and PXG10 were used to amplify exons 2, 3, 4, or the first 312-bp of exon 5, respectively (along with 30-50 bases, on average, of flanking intron. Each PCR reaction contained genomic DNA (0.1 μg), each of the designated primer combinations (10μM each), 10 mM Tris-HCl, pH 8.3 (at 25°C), 50 mM KCl, 1.5 mM MgCl, 0.001% (w/v) gelatin (Sigma, Cat. No. G2500, St. Louis, MO), 0.2 mM each dNTP and 2.5 units of Amplitaq polymerase. Reaction mixes were overlaid with mineral oil (50μl) and subjected to thirty cycles of 94°C for 1 min, 56°C for 1.5 min, and 72°C for 3 min, using a Perkin-Elmer Cetus thermocycler (Emeryville, CA). The final extension step at 72°C was for 30 min. Using normal DNA, these primer combinations result in the amplification of fragments which are 311, 164, 309, and 398-bp in length for exons 2 through 5, respectively.

PCR products were polished with Klenow, phosphorylated with T4 polynucleotide kinase, and subjected to electrophoresis through 1% low melting point agarose. Bands were visualized with ethidium bromide, excised, and blunt-end ligated with SmaI-cleaved and calf-intestinal alkaline phosphatase-dephosphorylated pBluescriptIISK+(0.1 μg) using T4 DNA ligase (1 unit). After incubation at 16°C for hr, the solidified agarose ligation reactions were melted, diluted ten fold with RF2 buffer (Hanahan, et al., In DNA Cloning Volume I: A Practical Approach, pp. 109-135 (1985)), and 50 μl was used to transform 100 μl of frozen competent E. coli strain XL1-Blue (Stratagene, La Jolla, CA) prepared using

Protocol #3 with the rubidium-containing buffers RF1 and RF2 (Hanahan, et al. DNA Cloning Volume I: A Practical Approach. pp. 109-135 (1985)). Ampicillin-resistant transformants were cultured in 7 ml of LB Broth which contained ampicillin (100 μg/ml), and DNA was purified and sequenced as previously described (Ritter et al., J. Biol. Chem., 265:7900-7906 (1990)). Each sample was sequenced in its entirety in both directions using the Bluescript primers KS and SK (Strategene, La Jolla, CA). Two or more of each subclone were sequenced.

Materials: Amplitaq polymerase was from Cetus (Emeryville, CA), T4 polynucleotide kinase from Bethesda Research Laboratories (BRL, Gaithersburg, MD), and the Large Klenow fragment from U.S. Biochemical (Cleveland, OH). All restriction endonucleases were from either New England Biolabs (Beverly, MA), Pharmacia-LKB (Piscataway, NJ), or BRL. Calf intestinal alkaline phosphatase was from Boehringer-Mannheim (Indianapolis, IN). The vector pBluescript was purchased from Stratagene.

Example II

Identification of a 13-Base Pair Deletion in Exon 2

The exons 2-5 (of Ex. I) were amplified individually using the polymerase chain reaction and subcloned as described in Example I. Analysis of exon 3, 4, and 5 subclones showed that each had the normal sequence (data not shown). However, a 13-bp deletion (Fig. 4) was observed in each of four independent exon-2 subclones. To confirm the mutation, four additional exon 2 subclones were

tested from a second PcR reaction, indicating that the mutation was not due to errors by Taq polymerase. The deletion is located 12-bp (downstream) from the intron 1/exon 2 junction (Fig. 4) and causes a change in the deduced amino acid sequence after residue 293 as seen in Fig. 4A. The dark vertical line by the nucleotides of the normal DNA sequences represents the deleted bases in the DNA of the Type I patient. The alterations suggest there should be no gross effects on the level of mRNA transcription, size, or translatability.

Example III

Determination of Whether More Than One UGT1 Locus Exists in the Genome

To establish whether more than one UGT1 locus exists in the genome, UGT1A-, UGT1D-, UGT1F-, and common exon-probes were used to analyze a set of Southern blots containing genomic DNA digested with three different enzymes predicted not to cleave exon 1 of UGT1A, UGT1F, or exons 2, 3, 4, or 5 (the common ones). A single band per enzyme digest for each probe, except for UGT1D, indicates that a single UGT1 complex exists in the genome (Fig. 6B). Since the UGT1D-probe revealed 2 to 4 more bands than predicted with EcoRI and BamHI digestions, respectively (Fig. 6B), it is suggested that 2 to 4 more UGT1D-like exon 1s exist further upstream. The UGT1 gene complex maps to chromosome 2 based on studies to localize the gene which encodes the HLUGP1 cDNA.

Example IV

Use of Agarose Gel Electrophoresis to Compare CN, Type I Patient Sample With Samples From Parents and That of a Normal Genome

Although it was established by a pedigree study that the parents of FB (the patient in question) are consanguineous, a demonstration of the pattern of inheritance by FB of the defective allele was carried out. Since CN, Type I is an autosomal recessive disease, it appeared that FB was probably homozygous for the same defective allele, UGT1*FB, and both parents were probably heterozygotic at the UGT1*FB locus. This hypothesis was confirmed using the polymerase chain reaction to amplify a fragment corresponding to the last 45-bp of intron 1 and first 154-bp of exon 2 followed by visualization of the reaction products with agarose gel electrophoresis (Fig. 5). Whereas a single band of the expected normal size (199-bp) was observed in a control sample from an unrelated normal individual (Lane 1), the sample corresponding to FB (Lane 3) contained a single band corresponding to a predicted fragment shorter by 13-bp (186-bp). These observations are consistent with FB being homozygous for the allele with a 13-bp deletion. In comparison, the mother (Lane 2) and father (Lane 4) exhibited both the 199-bp and 186-bp bands in equal amounts. Two other slower migrating bands not seen in either the normal or CN, Type I case were present in the parental samples. A sample representing a reaction which included a mixture of FB and normal DNA (Fig. 5, Lane 5), thereby simulating the composition of heterozygotic DNA, also exhibited

these sized bands. These observations are consistent with those from other PCR studies (Keen, et al., Trends Genet, 7:5 (1991)) with multiple alleles which suggest that the bands are due to mismatched DNA strands (i.e., heteroduplexes) which migrate differently from those which are perfectly matched.

Example V

Preparation of HUG-Br1 and HUG-Br2

Poly(A) RNA was prepared from human adult liver biopsy and converted to double-stranded cDNA using the Boehringer Mannheim cDNA synthesis kit and oligo (dT) as primer. The cDNA was ligated with EcoRI-SalI adaptors and cloned into the EcoRI site of lambda-ZAP bacteriophage (Stratagene, San Diego, CA). Unamplified phage (0.5 X 10^6 pfu) were screened with a 51-bp probe. In order to design a universal probe to detect all transferase cDNAs including that which is specific for bilirubin, all characterized (Mackenzie et al., J. Biol. Chem. 261, 6119-6125 (1986); Mackenzie et al., J. Biol. Chem. 261, 14112-14117 (1986); Harding et al., Proc.-Natl. Acad. Sci. U.S.A. 85, 8381-8385 (1988); Iyanagi et al., J. Biol. Chem. 261, 15607-15614 (1986); Jaye et al., Nucleic Acids Research 11, 2325-2335 (1983); Ritter et al., J. Biol. Chem. 265, 7900-7906 (1990)) transferase clones were analyzed for conserved sequences. It was uncovered that a 56-bp sequence is present near the carboxy terminus of all transferases examined. A 51-bp probe (5'-CAC GAC CTC ACC TGG TTC CAG TAC CAC TCT TTG GAT GTG ATT GGT

TTC CTG-3') corresponding to a 90% conserved motif (His-Asp-Leu-Thr-Trp-Phe-Gln-Tyr-His-Ser-Leu-Asp-Val-Ile-Gly-Phe-Leu) was used. This 17-amino acid sequence starting with a histidine residue at approximately position 481 is located in the lumen of the endoplasmic reticulum and overlaps the membrane-spanning region by 6 residues. It is speculated that this sequence, preceded by a conserved arginine at -4 residue, is a possible binding site for the common donor substrate, UDP-glucuronic acid. The probe was labeled to a specific activity of 0.22 X 10^9 dpm/ug by fill-in of overlapping sense and antisense oligonucleotides. The probe was hybridized as described (Jaye et al., Nucleic Acids Research 11, 2325-2335 (1983)) except that 45°C and 30°C were the hybridization and wash temperatures, respectively. All manipulations of lambda-ZAP and excision of the Bluescript plasmid containing positive inserts were performed as recommended by the manufacturer. Four clones representing independent copies of the HUG-Br1 cDNA and one clone of HUG-Br2 were isolated and characterized. The complete nucleotide sequence of both strands of Z17 and the longest of the HUG-Br1 inserts, Z6, were determined by double-stranded dideoxy sequencing in conjunction with deletions as described previously (Ritter et al., J. Biol. Chem. 265, 7900-7906 (1990)).

Complementary DNAs containing the remaining portion of the Z6 and Z17 coding and 5' untranslated regions were generated using the anchored PCR (RACE) method of Frohman et al. (PRC Protocols: A Guide to Methods and Applications pp.

28-38, Academic Press, New York (1990)). Antisense oligonucleotides (10 picomoles) corresponding to base positions +206 to +227 of the Z17-cDNA (5'-TGT GCC CTT AAA GTC TAA GTC A-3') or +235 to +251 (5'-CTG CAA GGA AGG AAA GG-3') of the Z6-cDNA inserts, respectively, were used to prime first-strand cDNA synthesis. After a tailing reaction with terminal deoxynucleotidyl transferase and dATP, the products were used as templates in separate PCR reactions with the adaptor-poly(dT) (Frohman et al., PRC Protocols: A Guide to Methods and Applications pp. 28-38, Academic Press, New York (1990) and a nested clone-specific oligonucleotide as primers. Major bands of approximately 470-bp (Z6-extended template) and 430-BP (Z17-extended template) were kinased and cloned into the ECORV site of Bluescript SK- (Stratagene, San Diego, CA), and individual clones were sequenced. Clones produced from duplicate PCR reaction had identical nucleic acid sequences suggesting that mutations were not introduced by TaqI polymerase during amplification, nor were forms with similar primer sequences co-amplified. The longest clones contained 449- and 399-bp inserts, respectively. The fragments extended Z6 by 234-bp and Z17 by 357-bp to complete the coding region with 15- and 29-bp, respectively, of 5'- untranslated sequence. Complete coding versions of the two new cDNAs, HUG-Br1 (original clone Z6) and HUG-Br2 (Z17) were assembled by ligating Z6 and Z17 to their respective 5'-end by using a unique restriction site near the 5'-terminus of each clone, AvaI (+298 of HUG-Br1) and XbaI (+360 of HUG-Br2). In both cases, the nucleotide sequence across the junction site was

verified by dideoxy sequencing. Both clones were judged to contain their bona fide 5' ends by virtue of the fact that a DNA fragment of the predicted size was successfully amplified from human liver cDNA using primers located at the 5' and 3' ends of the complete coding cDNAs. The nucleotide sequences of HUG-Br1 and HUG-Br2 have been submitted to the Genbank database. Further, E. coli transformed with the clones pHUG-Br1 and pHUG-Br2 were deposited in the American Type Culture Collection in Rockville, Maryland, U.S.A. under the Budapest Treaty on January 9, 1991. The clone pHUG-Br1 (deposited on January 9, 1991 under the name "E. coli derivative, XL-1 Blue transformed with pBluescriptSK+ HUG-Br1") has been assigned accession number 68510. The clone pHUG-Br2 (deposited on January 9, 1991 under the name "E.-coli derivative XL-1 Blue transformed with pBluescriptSK+ HUG-Br2") has been assigned accession number 68509.

Example VI

Northern Blot Analysis

Poly(A) 'RNA was isolated (Chirgqin et al., Biochemistry 18, 5294-5299 (1979)) from two normal liver biopsies (A.K. and W.E., adult females) and from the liver of a Type I Crigler-Najjar syndrome patient (female, age 3) who underwent transplant surgery (Kaufman et al., Heplatology 6: 1259-1262 (1986)). mRNA was also isolated from the livers of an untreated and a phenobarbital-treated Erythrocebus patas monkey maintained on 15 mg % in its drinking water for over one year. The mRNA was electrophoresed and blotted as described (Ritter, et

al., J. Biol. Chem. 265: 7900-7906 (1990)). The filters were hybridized with ^{32}p-labeled probes corresponding to the most 5' 725-bp EcoRI-fragment of HUG-Br1 and the most 5' 400-bp-fragment of HUG-Br2 (Church, et al., Proc. Nat. Acad. Sci. U.S.A. 81:1991-1995). The amounts of RNA applied to the gel were normalized (McKinnon, et al., Mol. Cell. Biol. 7: 2148-2154 (1987)) by hybridization to ^{32}p-human cyclophilin cDNA. All probes were radiolabeled by random primed synthesis using a kit (Pharmacia).

From this hybridization with 5-specific probes from HUG-Br1 and HUG-Br2 to Northern filters containing mRNA isolated from two normal individuals and a Type I Crigler-Najjar patient, it can be seen in Figure 10 (Panel A) that the size of both messages is 2.6kb. The relative abundance of the HUG-Br1 message is 3.0 fold higher than that for HUG-Br2. This determination is based on the amount of mRNA applied as normalized by hybridization to the ^{32}p-labeled-cyclophilin cDNA probe and the difference in the length of the 5' specific probes. By an analogous determination for the Crigler-Najjar patient, the data show that the HUG-Br2 message level is approximately normal but that for HUG-Br1 is depressed 4-fold. The fact that both mRNAs encoding HUG-Br1 and HUG-Br2 (although reduced in the case of HUG-Br1) are present in the Type I Crigler-Najjar patient and yet activity remains refractory to all treatments suggests that a subtle defect (e.g., point mutation, small insertion, or deletion) in the common region of a possibly shared

exon or a highly mutable region of a multigene model underlies the defect in Type I Crigler-Najjar.

To determine if either species of mRNA is regulated by phenobarbital, the present inventors compared the levels from the liver of an Old World monkey maintained on phenobarbital in the drinking water. The results show that HUG-Br2, the lower abundance mRNA, was induced 2-fold but not that for HUG-Brl (Fig. 10, Panel B) based on normalization to cyclophilin mRNA. The existence of two bilirubin transferase isozymes and the responsiveness of one to phenobarbital administration suggest that elevation in activity after this treatment is due to the increased expression of the low constitutive isoenzyme encoded by HUG-Br2. The results also suggest that the higher abundance mRNA encoding HUG-Rrl is most likely responsible for constitutive bilirubin activity and that this version may be underrepresented in both Gilbert's syndrome and in Type II Crigler-Najjar disease. By analogy, HUG-Br2-associated activity necessarily accounts for the elevation in activity after phenobarbital administration.

Example VII

Assay for Glucuronidation of ^{14}C-Bilirubin

For the transfection studies the expression units, pHUG-Brl and pHUG-Br2, contained HUG-Brl and HUG-Br2, respectively, in the correct orientation with respect to the promoter element of the pSVL plasmid, and a control unit contained HUG-Brl in the reversed orientation. Plasmid preparations and transfection studies were performed

as previously described (Ritter, et al., J. Biol. Chem. 265: 7900-7906). Cells were incubated 72 hour at 37°C, harvested, and stored at -70°C or assayed fresh.

Microsomes (2.9 mg) from 3-methylcholanthrene-treated CD1 mice and cell homogenates (0.8 or 1.9 mg) from transfected cells were assayed in a 200-uL reaction mixture containing 50mM Tris pH 7.7, 5mM MgCl, 20 mM UDP-glucuronic acid, 8.3 mM saccharic 1,4-lactone, and 25 uM ^{14}C-bilirubin (74,000 CPM, solubilized in 10 uL of dimethylsulfoxide) and incubated for 6.5 hour at 37°C. The sensitivity of the reaction product to ß-glucuronidase (where saccharic 1,4 lactone had been omitted) was then determined by adjusting the reaction to pH 7.0 with 1 M Na_2HPO_4, and adding 5 units of enzyme plus 3 mM ascorbic acid and continuing the incubation for 1 hour at 37°C. Microsomes and cell homogenates were preactivated with 0.5 mg of [3-(3-cholamidopropyl)-dimethylammonio]-1-propane (CHAPS) detergent to 1 mg protein before adding to the reaction mixture. All reagents used in the assay were made fresh in argonized double-distilled water, and all incubations were flushed with argon, sealed and incubated in the dark. In order to generate a more stable product for thin layer chromatography (TLC) analysis and to distinguish the forms of glucuronides, the ^{14}C-bilirubin mono- and diconjugates were selectively converted to the corresponding mono- and dimethyl esters by using alkali-catalyzed transesterification in the presence

of methanol and ascorbic acid as published (Blanckaert, N. (1980) Biochem. J. 185: 115-128; Hauser, et al., J. Biol. Chem. 259: 4527-4533 (1984)). Enriched bile was added as a carrier before the transesterification to confirm the location of radioactive product. The methyl esters of bilirubin corresponding to the IXαC8 and the IXoC12 monoglucuronides and the diglucuronide of bilirubin were extracted and chromatographed on TLC plates as described (Blanckaert, N. (1980) Biochem. J. 1985: 115-128; Hauser, et al., J. Biol. Chem. 259: 4527-4533 (1984)). Plates were analyzed for activity on an Ambis Radioanalytical Imaging Mark II System before exposure to x-ray film.

The assay for glucuuronidation of ^{14}C-bilirubin by mouse microsomes was carried out to establish the validity of the method as shown in Figure 11 (Panel A). Lanes 3 and 4 show, respectively, that the three bands attributed to IXαC8 and IXαC12 isomeric monoglucuronides (BMEs) and the diglucuronide (BDE) were present in a complete reaction mixture but were totally absent when UDP-glucuronic acid was omitted. The relative mobility of bilirubin and its methyl ester derivatives in this stem is Br > IXαC8 > IXαC12 > IXαC8, C12 according to the published method (Blanckaert, N. (1980) Biochem. J. 185: 115-128). The decrease (40%, as estimated by quantitation on AMBIS) in intensities of specific bands in Fig. 11 (Panel A) Lane 2 shows that the product was sensitive to ß-glucuronidase after one hour incubation. The band intensities also show that bilirubin diglucuronide was the primary conjugate in

mouse microsomes. Similarly, Lane 1 in Figure 11 (Panel A) shows that HUG-Br2- encoded activity in 1.9 mg of cell homogenate yielded the same three converted bilirubin methyl ester isomeric products in about equal amounts. In a separate experiment (Figure. 11 (Panel B)), HUG-Br1-encoded activity in 0.8 mg of cell homogenate also yielded the converted methyl ester derivatives in about equal amounts (Lane 2), and the glucuronie products, before conversion to the methyl esters, were sensitive to ß-glucuronidase (Lane 1). Cells mock-transfected with HUG-Br1 contained in pSVL in the reversed orientation with respect to promoter did not yield (Lane 3) methyl ester derivatives as seen in Figure 11 (Panel A) (Lanes 1, 2 and 3) and Figure 11 (Panel B) (Lane 2).

Example VIII

Sequence Homology

A comparison of the amino acid sequences of various UDP-glucuronosyltransferase forms: rat phenol UDP- glucuronosyltransferase (Rat PHENOL), human phenol UDP-glucuronosyltransferase (HLUGP1), human bilirubin UDP-glucuronosyltransferase, type I (HUG-Br1), human bilirubin UDP-glucuronosyl-transferase, type II (HUG-Br2), and rat bilirubin UDP-glucuronosyltransferase (Rat Bil), as set forth in Figure 12, shows that the carboxyl terminus of the various UDP-glucuronosyltransferase forms is conserved as between the human and the rat species and the phenol and bilirubin acceptor substrate forms. The human forms, HLUGP1, HUG-Br1 and HUG-Br2, show exact homology between amino acids

283-531, 285-533, and 296-534, respectively. The rat forms, Rat PHENOL and Rat Bil, show exact homology between amino acids 281-529 and 283-531, respectively. Further, there is a 90.6% conserved amino acid sequence between the two species.

In the nonconserved amino terminal regions of Rat PHENOL, HLUGP1, HUG-Br1, HUG-Br2 and Rat Bil, corresponding to amino acids 1-280, 1-282, 1-284, 1-285, and 1-282, respectively, there is 66% homology between HUG-Br1 and HUG-Br2, 70% homology between HUG-Br1 and Rat Bil, 76% homology between HUG-BR2 and Rat Bilo, 84% homology between Rat PHENOL and HLUGP 1, and 60% homology between Rat Bil and HLUGP 1. Comparisons of the rat bilirubin and phenol transferase amino acid sequences supports an association between the amino terminal sequences and substrate specificity. This region in rat bilirubin transferase is more similar to both of the two human bilirubin transferases (76 and 70%) than to human phenol transferase (60%). Similarly, the unique part of rat phenol transferase is much more highly related to the corresponding part of human phenol transferase (84%) than to either of the three bilirubin sequences (60-65%). The strict conservation shown in the carboxyl termini in either species suggests exon sharing within a single gene may occur [alternative 5' splicing as suggested (Sato, et al., Biochem. Biophys. Res. Commun. 169, 260-264 (1990)] to generate different isozymes in this transferase subfamily.

Example IX

Identification of HUG-Br3 from a Human Genomic Library Using a Probe Derived from HUG-Br2

The 470-bp 5' end of HUG-Br1 and the 450-bp 5' end of HUG Br2 were used as probes to screen a human genomic cosmid library. One clone was selected (approximately 40kb) which hybridized avidly to HUG-Br1 and very weakly to HUG-Br2 and 2 other clones (45 and 44kb) which hybridized to the unique region of HUG-Br2 and not to the unique region of HUG-Br1 or to the common fragment of HUG-Br1 and HUG-Br2 coding for amino acids 292-533. The three clones were determined by restriction enzyme analysis to overlap to generate a linear consensus fragment (97kb) (see Figure 13A and 13B). Eight subclones (3-14kb) were generated by digestion of the two 5' cosmids with restriction enzymes as indicated in Fig. 13A and 13B and ligating each fragment into similarly digested Bluescript plasmid. E. coli were transformed separately with each recombinant plasmid and plated on LB plates with ampicillin. Each transformed colony was grown in 10 mls of LB broth with ampicillin. Plasmid samples were prepared according to Ritter, et al., J. Biol. Chem. 265: 7900-7906 (1990). Plasmid DNA samples were analyzed by Southern blot analysis according to (Sambrook, Fritsch and Maniatis, Molecular Cloning-A-Laboratory Manual, Second Edition #2). Each one of the eight subclones was probed with the three fragments of HUG-Br2 (respectively, 1-450 bp; 454-666 bp; and 666-863 bp fragments representing the unique end of HUG-Br2). Four Southern blots, each containing DNA from each of the eight Bluescript

subclones, were made. Each of the HUG-Br2 fragments (defined above) was separately hybridized to one of the Southern blots and the fourth blot was hybridized to full-length HUG-Br1 to confirm that only the HUG-Br2 (1-863-bp) region hybridized to these subclones. Five separate clones were identified (1, 4, 6, 7, and 8), which hybridized to the 3 separate fragments of HUG-Br2. Other subclones were all introns. Each positive subclones was further mapped in detail by restriction enzymes. Each subclone was digested with two to three sets of two enzymes each and again analyzed by Southern blot analysis as already described.

Three sets of Southern blots were made with each containing the sets of double-digested samples. The three unique fragments of HUG-Br2 were again hybridized to a separate blot. The now smaller fragment (3-5kb) which contained the intact exon was subcloned into Bluescript as described above. Each subclone was transformed into E. coli and plasmid samples prepared as described. The recombinant plasmid was sequenced by making nested deletions with an Erase-a-Base kit (Promega, Madison, WI). Direct plasmid dideoxy sequencing was carried out on the deleted clones using the Sequel Ds/KF sequencing kit (IBI, New Haven, CT) according to the manufacturer's protocol and as described in the reference, Ritter, et al., J. Biol. Chem. 265: 7900-7906 (1990). The alignment of the HUG-Br3 sequence with the corresponding region of HUG-Br2 was carried out by the IBI Pustell sequence analysis software package. The preliminary data showing a high degree (> 90%) of homology between HUG-Br2 and

this short fragment of HUG-Br3 suggest that HUG-Br3 encodes a form of bilirubin UDP-glucurono-syltransferase.

One skilled in the art would recognize that the above-described methodology can be applied to uncover all mammalian bilirubin UDP-glucurono-syltransferases by optimizing hybridization stringency conditions and DNA probe length by routine techniques known to one skilled in the art.

Example X

Diagnostic Probes

The determination of the organization and the nucleotide sequence of the HUG-Br1 and HUG-Br2 gene complex has been completed using cosmid clones constructed from the genomic DNA of an individual not afflicted with defected in bilirubin glucuronidation. This nucleotide sequence information is considered to represent the normal genes. HUG-Br1, HUG-Br2, and the phenol (HLUGP1) clones have been isolated from a cDNA library constructed with mRNA isolated from the explanted liver of a Crigler Najjar Type I patient who underwent liver transplant surgery by using the unique region of each clone as a probe. The full-length cDNAs were sequenced using the direct plasmid method already described. The nucleotide sequence of the clones derived from the Crigler-Najjar Type I patient were compared to the sequences of the normal HUG-Br1, HUG-Br2 and HLUGP1 to determine if there are altered bases, deletions or insertions. The comparisons were made using the IBI Pustell Sequence software computer package.

The location of a modified nucleotide sequence was assessed at the level of the gene in the same patient's genomic DNA. Opposite strand primers (5--3') were synthesized with an 18 based complementary sequence to HUG-Br1, HUG-Br2, and HLGUP1 at positions 100 bases upstream and 100 bases downstream of the altered nucleotides. Using the specific primers, sufficient quantities of the target DNA from normal genome and from the defective genome were generated by PCR using a Perkin-Elmer Kit (Norwalk, CT). The two types of amplified DNA as well as samples of the HUG-Br1, HUG-Br2 and HLUGP1 cDNA previously isolated above from the Crigler-Najjar Type I cDNA library and the normal HUG-Br1 and HUG-Br2 cDNA clones (as heretofore described) were analyzed by Southern blot analysis. Each blot was made in duplicate. An 18-base oligo reflecting the altered DNA sequence(s) in the Crigler-Najjar Type I DNA, as well as an 18-based oligo reflecting this sequence in DNA from normal genome were made. Each 18-base oligonucleotide probe was hybridized to the Southern blot prepared above using high stringency conditions described in Sambrook, Fritsch and Maniatis Molecular Cloning - A Laboratory Manual. One skilled in the art would recognize that the normal 18-base probe will hybridize only to the mutated cDNAs and the mutated PCR amplified 200-base target sequence.

Genomic DNA from patients with Gilbert's disease and Crigler-Najjar Type II syndrome can be analyzed for altered DNA sequence(s) compared to normal. The DNA will be analyzed for alterations by first making PCR products to exonic regions of the

HUG-Br1 and HUG-Br2 and other bilirubin genes as yet to be determined (e.g., HUG-Br3). Plasmid subclones of the PCR products will be made and sequenced. Upon making comparisons to normal DNA any altered sequences will be subjected to the same analyses already described for Crigler-Najjar Type I as above.

The validity of the diagnostic probe is established by the selective hybridization under high stringency conditions) of each probe to target DNA containing a perfect match.

Whereas the clones according to the invention have utilized as diagnostic probes or for use in a diagnostic assay as described above, said clones would also have utility in gene therapy techniques for Crigler-Najjar, Type I disease.

* * * * *

While the invention has been described with reference to certain specific embodiments, it will be appreciated that many modifications and changes may be made by those skilled in the art without departing from the spirit of the invention. It is intended, therefore, by the appended claims, to cover all such modifications and changes as fall within the true scope of the invention.

CLAIMS:

1) An isolated gene locus, referred to as UGT1, comprising the nucleotide sequence shown in Figure 1.

2) A DNA segment selected from the group consisting of exon 1 of UGT1A, UGT1BP, UGT1C, UGT1D, UGT1E, and UGT1F.

3) A recombinant DNA molecule comprising a vector and said DNA segment of claim 2 referred to exon 1 of UGT1A.

4) A recombinant DNA molecule comprising a vector and said DNA segment of claim 2 referred to as exon 1 of UGT1BP.

5) A recombinant DNA molecule comprising a vector and said DNA segment of claim 2 referred to as exon 1 of UGT1C.

6) A recombinant DNA molecule comprising a vector and said DNA segment of claim 2 referred to exon 1 as UGT1D.

7) A recombinant DNA molecule comprising a vector and said DNA segment of claim 2 referred to exon 1 as UGT1E.

8) A recombinant DNA molecule comprising a vector and said DNA segment of claim 2 referred to as exon 1 of UGT1F.

9) A DNA segment that codes for a polypeptide having an amino acid sequence corresponding to a mammalian UDP-glucurono-syltransferase form or a mutant or variant form thereof.

10) A recombinant molecule comprising a vector and said DNA segment of claim 9.

11) A DNA segment that codes for a polypeptide having an amino acid sequence corresponding to human bilirubin UDP-glucuronosyltransferase type I.

12) A DNA segment that codes for a polypeptide having an amino acid sequence corresponding to human bilirubin UDP-glucuronosyltransferase type II.

13) The human liver cDNA segment HUGBr1, ATCC accession under 68510.

14) The human liver cDNA segment HUGBr2, ATCC accession under 68509.

15) A recombinant DNA molecule comprising a vector and said DNA segment of claim 11.

16) A recombinant DNA molecule comprising a vector and said DNA segment of claim 12.

17) A recombinant DNA molecule comprising a vector and said human liver cDNA segment HUG-Br1 of claim 13.

18) A recombinant DNA molecule comprising a vector and said human liver cDNA segment HUG-Br2 of claim 14.

19) A prokaryotic cell transformed with said recombinant DNA molecule of claim 10.

20) A prokaryotic cell transformed with said recombinant DNA molecule of claim 15.

21) A prokaryotic cell transformed with said recombinant DNA molecule of claim 16.

22) A prokaryotic cell transformed with said recombinant DNA molecule of claim 17.

23) A prokaryotic cell transformed with said recombinant DNA molecule of claim 18.

24) A eukaryotic cell transfected with said recombinant DNA molecule of claim 10.

25) A eukaryotic cell transfected with said recombinant DNA molecule of claim 15.

26) A eukaryotic cell transfected with said recombinant DNA molecule of claim 16.

27) A eukaryotic cell transfected with said recombinant DNA molecule of claim 17.

28) A eukaryotic cell transfected with said recombinant DNA molecule of claim 18.

29) A diagnostic probe of at least 18 bases for the diagnosis of a syndrome selected from the group consisting of Gilbert's syndrome, Crigler-Najjar syndrome Type I, and Crigler-Najjar syndrome Type II comprising a specific sequence of a cDNA encoding a mammalian bilirubin UDP-glucurono-syltransferase enzyme, mutant form, or variant thereof, said diagnostic probe identifying the corresponding mammalian genomic DNA.

30) A diagnostic probe of at least 18 bases for the diagnosis of a syndrome selected from the group consisting of Gilbert's syndrome, Crigler-Najjar syndrome Type I and Crigler-Najjar syndrome Type II comprising a DNA sequence having sufficient homology to a cDNA encoding a mammalian bilirubin UDP-glucuronosyltransferase form, mutant form, or variant thereof to identify the mammalian genomic DNA corresponding to said cDNA.

31) A diagnostic probe for the detection of Crigler-Najjar, Type I syndrome, wherein said probe comprises a DNA sequence corresponding to the 13-bp nucleotide deletion present in exon 2 of said DNA segment of claim 9.

32) A PCR primer pair designed to amplify a specific portion of mammalian genomic DNA, wherein each member of said primer pair is of from 17 to 20 bases, and further wherein each member of said primer pair comprises a specific sequence of a cDNA encoding a mammalian bilirubin UDP-glucurono-syltransferase form, mutant form, or variant thereof.

33) A diagnostic assay for a syndrome selected from the group consisting of Gilbert's disease, Crigler-Najjar syndrome Type I and Crigler-Najjar syndrome Type II comprising the steps of:

(a) amplifying a sequence of mammalian genomic DNA using a PCR primer pair according to claim 32 to produce multiple DNA copies of said mammalian genomic DNA, and

(b) exposing said amplified DNA according to (a) the diagnostic probe of claims 29 or 30 under hybridization conditions to identify a specific mammalian genomic sequence characteristic of Gilbert's disease, Crigler-Najjar syndrome Type I or Crigler-Najjar syndrome Type II.

34) A diagnostic assay for the detection of Crigler-Najjar, Type I syndrome in a patient comprising the steps of:

a) altering one of a pair of normal PCR primers such that said altered primer contains the deletion mutation present in exon 2 of UGT1F of the genome of Crigler-Najjar, Type I patients;

b) adding said primers to a DNA sample of a patient suspecting of having said syndrome; and

c) determining whether hybridization occurs such that a PCR product results, lack of said product indicating presence of Crigler-Najjar, Type I syndrome.

35) An amino acid sequence as shown in Figure 9 wherein said sequence corresponds to human bilirubin UDP-glucuronosyltransferase, Type I (HUG-Br1).

36) An amino acid sequence as shown in Figure 9 wherein said sequence corresponds to human bilirubin UDP-glucuronosyltransferase, Type II (HUG-Br2).

SUBSTITUTE SHEET

FIG. IA

```
tgacacggccatagttggttcatattaaccatgtgattaaaatggttaaatattaatttgggttcttacatatcaaagggtaaaattcagagcaag

ggagaGGTAGACAGGACCTGTGAAAAGCAGTGGTTAGTTTAGGGAAAATACCTAGGAGCCCTGTGATTTGGAGAGTGAAAACTCTTTATTACCGTT

GTTACTTTAACTCTTTCCAGGATGGCCTGCCTCCTTCGCTCATTTCAGAGAATTTCTGCAGGGGTTTTCTTCTTAGCACTTTGGGGCATGGTTGTA
                      M  A  C  L  L  R  S  F  Q  R  I  S  A  G  V  F  F  L  A  L  W  G  M  V  V

GGTGACAAGCTGCTGGTGGTCCCTCAGGACGGAAGCCACTGGCTTAGTATGAAGGATATAGTTGAGGTTCTCAGTGACCGGGGTCATGAGATTGTA
 G  D  K  L  L  V  V  P  Q  D  G  S  G  W  L  S  M  K  D  I  V  E  V  L  S  D  R  G  H  E  I  V

GTGGTGGTGCCTGAAGTTAATTTGCTTTTGAAAGAATCCAAATACTACACAAGAAAAATCTATCCAGTGCCGTATGACCAAGAAGAGCTGAAGAAC
V  V  V  P  E  V  N  L  L  L  K  E  S  K  Y  Y  T  R  K  I  Y  P  V  P  Y  D  Q  E  E  L  K  N

CGTTACCAATCATTTGGAAACAATCACTTTGCTGAGCGATCATTCCTAACTGCTCCTCAGACAGAGTACAGGAATAACATGATTGTTATTGGCCTG
 R  Y  Q  S  F  G  N  N  H  F  A  E  R  S  F  L  T  A  P  Q  T  E  Y  R  N  N  M  I  V  I  G  L

ATACTTCATCAACTGCCAGAGCCTCCTGCAGGACAGGGACACCCTGAACTTCTTTAAGGAGAGCAAGTTTGATGCTCTTTTCACAGACCCAGCCTT
  Y  F  I  N  C  Q  S  L  L  Q  D  R  D  T  L  N  F  F  K  E  S  K  F  D  A  L  F  T  D  P  A  L

ACCCTGTGGGGTGATCCTGGCYGAGTATTTGGGCCTACCATCTGTGTACCTCTTCAGGGGTTTTCCGTGTTCCCTGGAGCATACATTCAGCAGAAG
  P  C  G  V  I  L  A  E  Y  L  G  L  P  S  V  Y  L  F  R  G  F  P  C  S  L  E  H  T  F  S  R  S

CCCAGACCCTGTGTCTCACATTCCCAGGTGCTACACAAAGTTTTCAGACCACATGACTTTTTCCCAACGAGTGGCCAACTTCCTTGTTAATTTGT
  P  D  P  V  S  Y  I  P  R  C  Y  T  K  F  S  D  H  M  T  F  S  Q  R  V  A  N  F  L  V  N  L

TGGAGCCCTATCTATTTTATTGTCTGTTTTCAAAGTATGAAGAACTCGCATCAGCTGTCCTCAAGAGAGATGTGGATATAATCACCTTATATCATG
L  E  P  Y  L  f  Y  C  L  F  S  K  Y  E  E  L  A  S  A  V  L  K  R  D  V  D  I  I  T  L  Y  Q

AAGGTCTCTGTTTGGCTGTTAAGATATGACTTTGTGCTTGAATATCCTAGGCCGGTCATGCCCAACATGGTCTTCATTGGAGGTATCAACTGTAAGA
 K  V  S  V  W  L  L  R  Y  D  F  V  L  E  Y  P  R  P  V  M  P  N  M  V  F  I  G  G  I  N  C  K

AGAGGAAAGACTTGTCTCAGgttggt
K  R  K  D  L  S  Q
```

UGTIF Exon 1

FIG. IA (CON'T)

gggtttatttctttggctgccttgtttcttccaggctctgtcctccctcactcatttggctccttagagccgactgtcccttggaggatttcctg

agaacggtgggggaagtgatacccgctcggagcagcgggaa<——15.3 kbp——>caattacatgttgattgcgaggtctcagtgacaagg

taattaagacgaaggaaacaattctaggaggCACAACGTGGGGT

FIG. IB

UGTIE Exon I

```
GGACAGTCAGCTGTCGGTGGCTTCTGCTGAGATGGCCACAGGACTCCAGGTTCCCCTGCCGCAGCTGGCCACAGGACTGCTGCTTCTCCTCAGTGT
                    M  A  T  G  L  Q  V  P  L  P  Q  L  A  T  G  L  L  L  L  L  S  V
CCAGCCCTGGGCTGAGAGTGGGAAGGTGCTGGTGGTGCCCACTGATGGCAGCCACTGGCTCAGCATGCGGGAGGCCTTGCGGGACCTCCATGCGAG
  Q  P  W  A  E  S  G  K  V  L  V  V  P  T  D  G  S  H  W  L  S  M  R  E  A  L  R  D  L  H  A  R

AGGCCACCAGGTGGTGGTCCTCACCCTGGAGGTGAATATGTACATCAAAGAAGAGAACTTTTTCACCCTGACAACGTATGCCATTTCATGGACCCA
  G  H  Q  V  V  V  L  T  L  E  V  N  M  Y  I  K  E  E  N  F  F  T  L  T  T  Y  A  I  S  W  T  Q

GGACGAATTTGATCGCCTTTTGCTGGGTCACACTCAATCGTTCTTTGAAACAGAACATCTTCTGATGAAATTTTCTAGAAGAATGGCAATTATGAA
  D  F  F  D  R  L  L  L  G  H  T  Q  S  F  F  E  T  E  H  L  L  M  K  F  S  R  R  M  A  I  M  N

CAATATGTCTTTGATCATACATAGGTCTTGTGTGGAGCTACTGCATAATGAGGCCCTGATCAGGCACCTGCATGCTACTTCCTTTGATGTGGTTC
  N  M  S  L  I  I  H  R  S  C  V  E  L  L  H  N  E  A  L  I  R  H  L  H  A  T  S  F  D  V  V

TAACAGACCCCTTTCACCTCTGCGCGGTGCTGGCTAAGTACCTGTCGATTCCTGCTGTGTTTTCTTGAGGAACATTCCATGTGATTTAGA
L  T  D  P  F  H  L  C  A  A  V  L  A  K  Y  L  S  I  P  A  V  F  F  L  R  N  I  P  C  D  L  D

CTTTAAGGGCACACAGTGTCCAAACCCTTCCTCCTATATTCCTAGATTACTAACGACCAATTCAGACCACATGACATTCCTGCAAAGGGTCAAGAA
  F  K  G  T  Q  C  P  N  P  S  S  Y  I  P  R  L  L  T  T  N  S  D  H  M  T  F  L  Q  R  V  K  N

CATGCTCTACCCTCTGGCCCTGTCCTACCTTTGCCATGCTGTTTCTGCTCCTTATGCAAGCCTTGCCTCTGAGCTTTTTCAGAGAGAGGTGTCAGT
  M  L  Y  P  L  A  L  S  Y  L  C  H  A  V  S  A  P  Y  A  S  L  A  S  E  L  F  Q  R  E  V  S  V

GGTGGAGCTTCTCAGCCATGCATCCTGGTGGCTGTTCCGAGGGGAGTTTGTGATGGATTACCCCAGGCCGATCATGCCCAACATGGTCTTCATTGG
  V  D  L  V  S  H  A  S  V  W  L  F  R  G  D  F  V  M  D  Y  P  P  I  M  P  N  M  V  F  I  G

GGGCATCAACTGTGCCAACGGGAAGCCACTATCTCAGgtctgtattggtgcctttatccaatcaatg<——7.5 kbp——>gtcagatgagctt
  G  I  N  C  A  N  G  K  P  L  S  Q

tttt

caagataggcgtgattggtctttcccaggttgggcccataacgaaaggcagttatacattaatgggtaataagtaactggaggagggcactttgt

cttccaattacatgctgatttgct
```

FIG. 1C

UGTID Exon I

aggtggctcaatgacaaggtaattaaggcgaaggaaacaaatgtagcaggCACAGCGTGGGGTGGACAGTCAGCTGTCGGTGGCTTCTGCTGAGAT
M

GGCCAGAGGACTCCAGGTTCCCCTGCCGCGGCTGGCCACAGGACTGCTGCTCCTCCTCAGTCTCCAGGCCCGGGCTGAGAGTGGAAAGGTGTTGGT
A R G L Q V P L P R L A T G L L L L L L S V Q P W A E S G K V L V

GGTGCCCACTGATGGCAGCCCCTGGCTCAGCATGCGGGAGGCCTTGCGGGAGCTCCATGCCAGAGGCCACCAGGCGGTGGTCCTCACCCCAGAG
V P T D G S P W L S M R E A L R E L H A R G H Q A V V L T P E

GTGAATATGCACATCAAAGAAGAGAAATTTTTCACCCTGACAGCCTATGCTGTTCCATGGACCCAGAAGGAATTGGAGCTACTGCATAATGAGGCC
V N M H I K E E K F F T L T A Y A V P W T Q K E F E L L H N E A

CTGATCAGGCACCTGAATGCTACTTCCTTTGATGTGGTTTTAACAGACCCCGTTAACCTCTGTGGGGCGGTGCTGGCTAAGTACCTGTCGATTCCT
L I R H L N A T S F D V V L T D P V N L C G A V L A K Y L S I P

GCTGTGTTTTTTTGGAGGTACATTCCATGTGACTTAGACTTTAAGGCCATACAGTGTCCAAATCCTTCCTCCTATATTCCTAAGTTACTAACGACCA
A V F F W R Y I P C D L D F K G T Q C P N P S S Y I P K L L T T

AITCAGACCACATGACATTCCTGCAAAGGGTCAAGAACATGCTCTACCCTCTGGCCCTGTCCTACATTTGCCATACTTTTTCTGCCCCTTATGCAA
N S D H M T F L Q R V K N M L Y P L A L S Y I C H T F S A P Y A

GTCTTGCCTCTGAGCTTTTTCAGAGAGAGGTGTCAGTGGTGGATCTTGTCAGCTATGCATCCGTCTGGCTGTTCCGAGGGGACTTTGTGATGGACT
S L A S E L F Q R E V S V V D L V S T A S V W L F R G D F V M D

ACCCCAGGCCGATCATGCCCAACATGGTCTTCATTGGGGGCATCAACTGTGCCAACGGGAAGCCACTATCTCAGgtctgtattggtgccttcatccY
P R P I M P N M V F I G G I N C A N G K P L S Q

aatcaatgttccaggcaaaacacttttaaaaaatgtatttacttacaagtgcttccatatctactttccaaag<——9.7 kbp——>a

gggcactctgtctttccaattacacgttgatttgctaagtggctcagtgacaaggtaattaagatgaagaaagcaaatgtagcaggCACAGCGTGGG

GTGGACAGTCAGCTGTCCGTGTCT

FIG. 1D

UGTIC Exon I

```
TCTGCTGAGATGGCCACAGGACTCCAGGTTCCCCTGCCGTCGCTGGCCACAGGACTGCTGCTTCTCCTCAGTGTCCAGCCCTGGGCTGAGAGTGGA
                M  A  T  G  L  Q  V  P  L  P  W  L  A  T  G  L  L  L  L  L  S  V  P  P  W  A  E  S  G

AAGGTGTTGGTGGTGCCCATTCATGGCAGCCACTGGCTCAGCATGCGGGAGGTCTTGCGGGAGCTCCATGCCAGAGGCCACCAGGCAGTGGTCCTC
  K  V  L  V  V  P  I  D  G  S  G  W  L  S  M  R  E  V  L  R  E  L  H  A  R  G  H  Q  A  V  V  L

ACCCCAGAGGTGAATATGCACATCAAAGAAGAGAACTTTTTCACCCTGACAACCTATGCCATTTCGTGGACCCAGGATGAATTTGATCGCCATGTG
  T  P  E  V  N  M  H  I  K  E  E  N  F  F  T  L  T  T  Y  A  I  S  W  T  Q  D  E  F  D  R  H  V

CTGGGCCACACTCAACTGTACTTTGAAACAGAACATTTTTCTGAAGAAATTTTTCAGAAGTATGGCAATGTTGAACAATATGTCTTTGGTCTATCAT
  L  G  H  T  Q  L  Y  F  E  T  E  H  F  L  K  K  F  F  R  S  M  A  M  L  N  N  M  S  L  V  Y  H

AGGTCTTGTGTGGAGCTACTACATAATGAGGCCCTGATCAGGCACCTGAATGCTACTTCCTTTGATGTGGTTTTAACAGACCCCGTTAACCTCTGC
  R  S  C  V  E  L  L  H  N  E  A  L  I  R  H  L  N  A  T  S  F  D  V  V  L  T  D  P  V  N  L  C

GCGGCAGTGCTGGCTAAGTACCTGTCGATTCCTACTGTGTTTTTTTTGAGGAACATTCCATGTGATTTAGACTTTAAGGGCACACAGTGTCCAAAC
  A  A  V  L  A  K  Y  L  S  I  P  T  V  F  F  L  R  N  I  P  C  D  L  D  F  K  G  T  Q  C  P  N

CCTTCCTCCTATATTCCTAGATTACTAACAACCAATTCAGACCACATGACATTCATGCAAAGGGTCAAGAACATGCTCTACCCTCTGGCCCTGTCC
  P  S  S  Y  I  P  R  L  L  T  T  N  S  D  H  M  T  F  M  Q  R  V  K  N  M  L  Y  P  L  A  L  S

TACATTTGCCATGCTTTTTCTGCTCCTTATGCCAGCCTTGCCTCTGAGCTTTTTCAGAGAGAGGTGTCAGTGGTGGATATTCTCAGTCATGCATCT
  Y  I  C  H  A  F  S  A  P  Y  A  S  L  A  S  E  L  F  Q  R  E  V  S  V  V  D  I  L  S  H  A  S

GTGTGGCYGTTCCGAGGGGACTTTGTGATGGACTACCCCAGGCCAATCATGCCCAACATGGTCTTCATTGGGGCATCAACTGTGCCAACAGGAAGC
  V  W  L  F  R  G  D  F  V  M  D  Y  P  R  P  I  M  P  N  M  V  F  I  G  G  I  N  C  A  N  R  K

CACTATCTCAGgtctgtattggtgccttcatccaatcaatygttccaggcaaaacattttaaaaaatgtatttattaa<——17.0 kbp——
P  L  S  Q
```

FIG. IE

UGTIBP Exon I

```
---tgatgatttgctaagtggctcagtgacaaggtaattaagatgaagaaacaaatgtaggaggCACAGCGTGGGGTGGACAGTCAGCTGTCCGT

GTCTTCTGCTGAGATGGCCACAGGACCCCAGGTTCCCCTGCTGTGGCTGGCCACAGGAGTGCTACTCCTTCTCAGTGCCCAGCCCTGGGCTGAGAGT
                M  A  T  G  P  P  Q  V  P  L  L  W  L  A  T  G  V  L  L  L  L  S  A  Q  P  W  A  E  S

GGGAAGGTGCTGGTGGTACCCATTGATGGCAGCCACTGGCTCGCATGCGGGAGGCCGTGCGGGAGGCTCCATCTCAGAGGCCACCAGGCGGTGGTCC
G  K  V  L  V  V  P  I  D  G  S  H  W  L  A  C  G  R  P  C  G  S  S  M  P  E  A  T  R  R  W  S

TCACCCCAGAGGTGAATATGCACATCAAACAAGAGAAATTTTTCACCCTGACAACCTATGCCATTTCATGGACCCAGGATGAATTTGATCGCCTTG
S  P  Q  R

TGCTGGGCCACACTCAACTGTACTTTGAAACAGAACATTTTCTGAAGACATTTTCTAAAAGTATGGCAATTTTGAAAAATTCATCTTTGGTCTTTC

ATAGGTCTTGTGTGGAGCTACTGCATAATGAGGCCCTGATCAGGCACCTGAATGCTACTTCCTTCGATGTGGTTTTAATAGACCCAATTTACCTCT

GTGGGGCAGTGGCTAAGTACCTATCAATTCCTGCTGTGTTTTTTTTTGAGGAACATTCCATGTGATTTAGACTTTAAGGGCACACAGTGTCCAAATC

CTTCCTCCTATATTCCTAAGTTACTAACGACCAATTCAGACCACATGACATTCCTGCAAAGGGTCAAGAACATGCTCTACCCTCTGGCCCTGTCCT

ACCTTTGCCATGCTCTTTCTGCTCCTTATGCAAGCCTTGCCTCTGAGCTTTTTCAGAGGGAGGTGTCAGTGGTGGATCTTCTCAGCCATGCATCTG

TGTGGCTGTTCCGAGGGGACTTTGTGATGGATTACCCCAGGCCGATCATGCCCAACATGGTCTTCATTGGGGGCATCAACTGTGCCAACAGGAAGC

CACTATCTCAGgtcggtgttcgtgccttcatccaatgttccaggcaaaacagtttttaaaaattgtatttacttacaattgcttccatatct

acttatctttctggtgcttt

tatttrtgcttgtcgtaatagccttcagtgagataaactgttaagggtctctagtagcgtatttcaggttttaaatggtcaatgagaggaagaag

aggcagacatgaatgtttatcaaaggatggactaggactgatgtgactcatggagatgttcgtttgtaaaggcaccatcttcatggttgtgcatgt

ccttcattgggaaggagcagggacactacattgagacttgaccatctggacttttttgtttgcacaatttttcaggggaaagatgatgcaacagcaa
```

SUBSTITUTE SHEET

UGTIA Exon I

FIG. IF

attacaattgttaatgtgataatttttagtggtcctgtctcttgcgaatgatagaggaggtgaccacaggggg

ggcccggtacc<----11.7 kbp ----><u>attaacttggtgtatcgattggttttgccatatatatataagtaggagagggcgaacctct</u>

ggCAGGAGCAAAGGCGCCATGGCTGTGGAGTCCCAGGGCGGACGCCCACTTGTCCTGGGCCTGCTGCTGTGTGCTGGGCCCACTGGTGTCCCAT
M A V E S Q G G R P L V L G L L L C V L G P V V S H

GCTGGGAAGATACTGTTGATCCCAGTGGATGGCAGCCACTGGCTGAGCATGCTTGGGGCCATCCAGCACTTGCAGCAGAGGGGACATGAAATTGTC
A G K I L L I P V D G S H W L S M L G A I Q Q L Q Q R G H I V V

CTAGCACCTGACGCCTCGTTGTACATCAGAGACGGAGCATTTTACACCTTGAAGACGTACCCTGTGCCAATCCAAAGGGAGGATGTGAAAGAGTCT
L A P D A S L Y I R D G A F Y T L K T Y P V P F Q R E D V K E S

TTTGTTAGTCTCGGGCATAATGTTTTTTGAGAATGATTCTTTCCTGCAGCGTGTGATCAAAACATACAAGAAAATAAAAAAGGACTCTGCTATGCT
F V S L G H N V F E N D S F L Q R V I K T Y K K I K K D S A M L

CCTGTCTCTGCCCACTGTATTCTTCTTGCATGCACTGCCAGCCTGGAATTTGAGGCTACCCAGTGCCCCAACCCATTCTCCTACGTGCCCAG
I S L P T V F F L H A L P C S L E F E A T Q C P N P F S Y V P R

GCCTCTCTCCTCTCATTCAGATCACATGACCTTCCTGCAGCGGGTGAAGAACATGCTCATTGCCTTTTCACAGAACTTTCTGTGGCGACGTGGTTTA
P L S S H S D H M T F L Q R V K N M L I A F S Q N F L C D V V Y

AACCCTTGCCTCAGAATTCCTTCAGAGAGAGGTGACTGTCCAGGACCTATTGAGCTCTGCATCTGTCTGGCTGTTTAGAAGTGACTTTGTGAAGGA
I L A S E F L Q R F V T V Q D L L S S A S V W L F R S D F V K D

TGACCCTAGGCCCATCATGCCCAATATGGTTTTTGTTGGTGGAATCAACTGCCTTCACCQAAATCCACTATCCCAGgtgtgtattggagtgggact
Y P R P I M P N M V F V G G I N C L H Q N P L S Q

tttaratgcgtatatttctttcagatgtattaactttggatcgattaactagccccatatgctgagcaagcattctgagata

atttaaaatgccctctttgttaatttttgactcctaggtttgagtctgtctttggcatcatcttctggatgatttcttgtatctgagatttcrgg

gtttagcatttcctttggtcatttta

FIG. IG

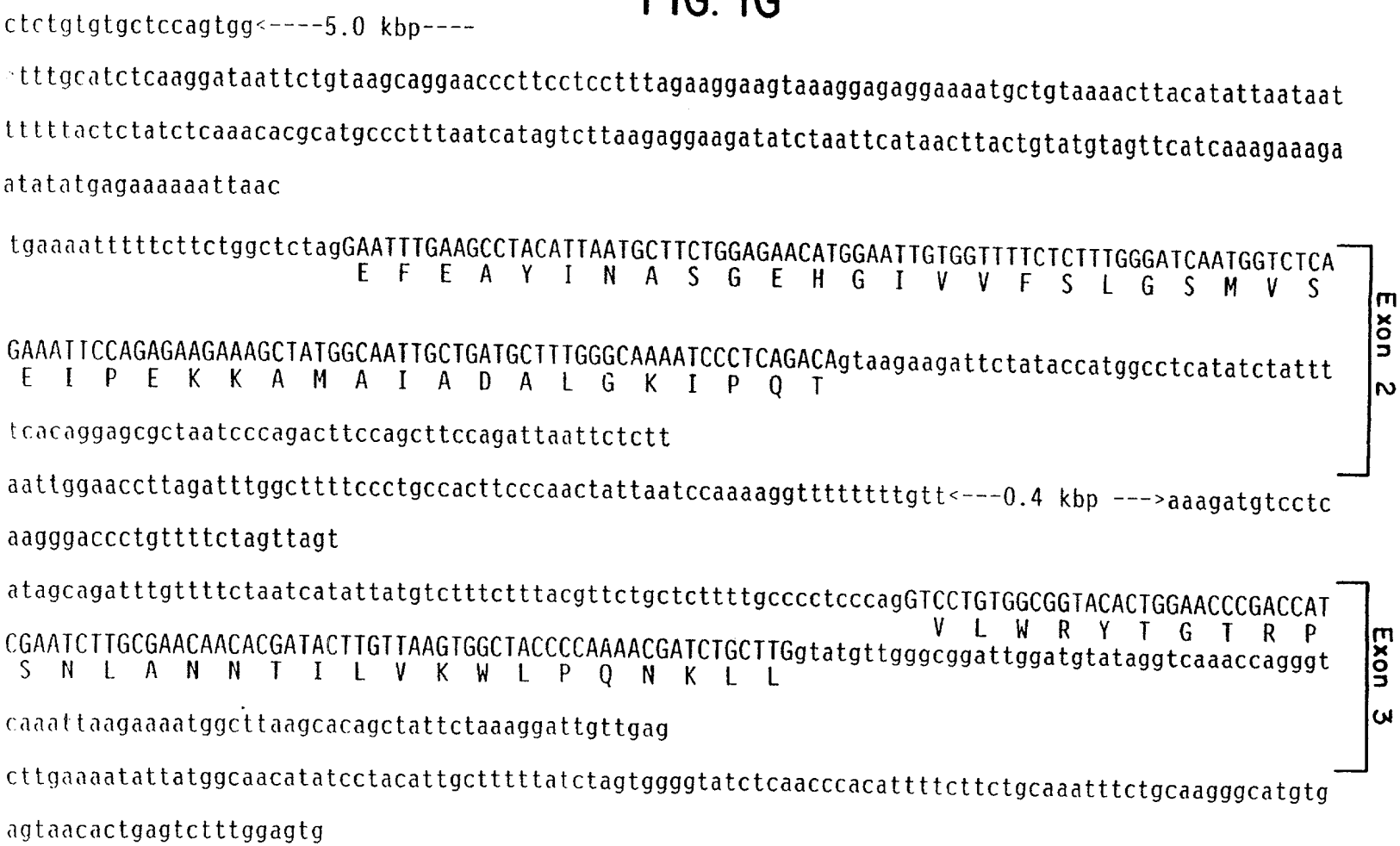
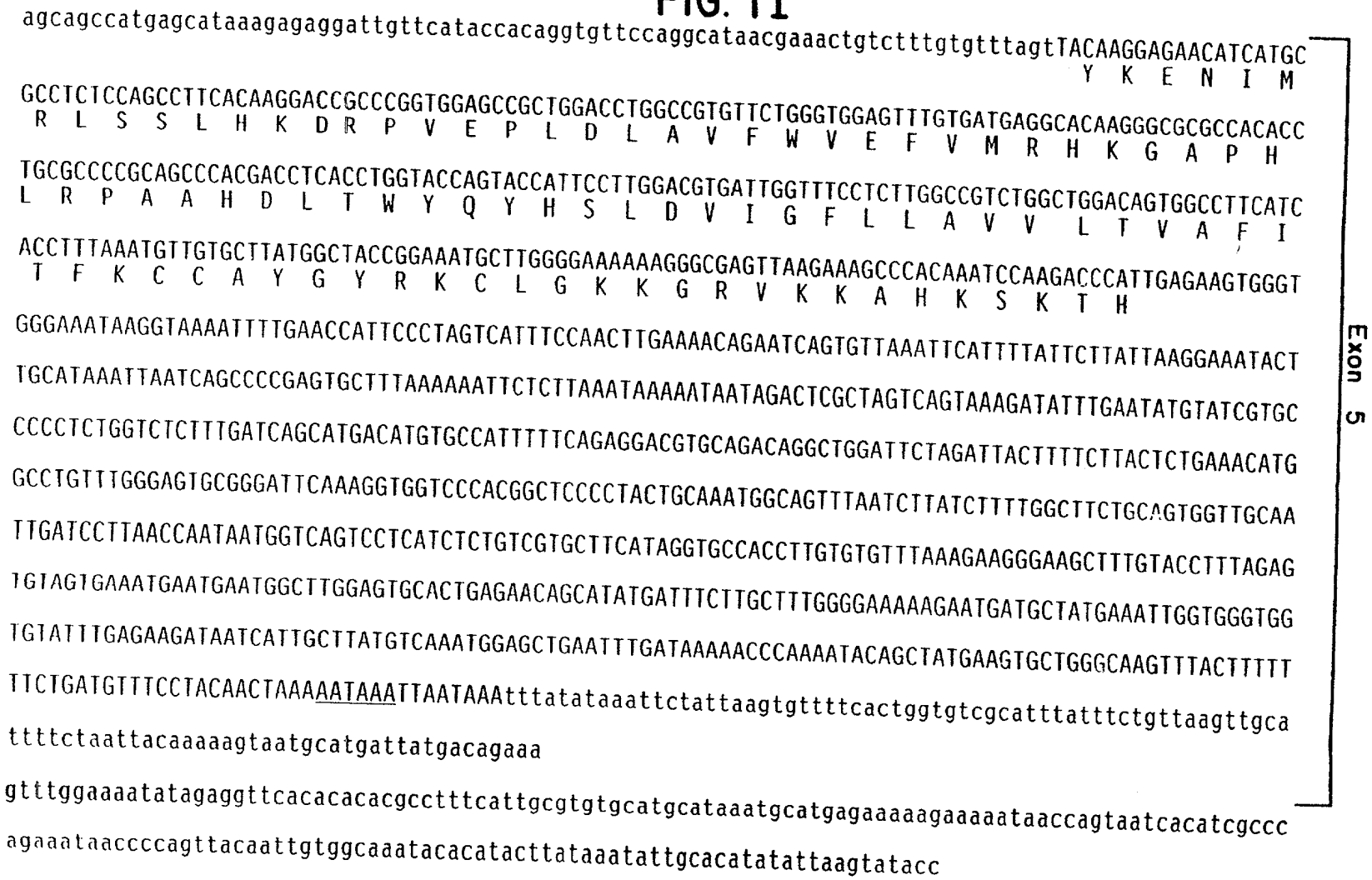


FIG. 1H

Exon 4

```
ttttcagaacctagatgtgtccagctgtgaaactcagagatgtaactgctgacatcctccctatttgcatctcagGTCACCCGATGACCCGTGCC
                                                                          G  H  P  M  T  R  A

TTTATCACCCATGCTGGTTCCCATGGTGTTTATGAAAGCATATGCAATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAGATGGACAAT
 F  I  T  H  A  G  S  H  G  V  Y  E  S  I  C  N  G  V  P  M  V  M  M  P  L  F  G  D  Q  M  D  N

GCAAAGCGCATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTGGAAATGACTTCTGAAGATTTAGAAAATGCTCTAAAAGCAGTCATCAAT
 A  K  R  M  E  T  K  G  A  G  V  T  L  N  V  L  E  M  T  S  E  D  L  E  N  A  L  K  A  V  I  N

GACAAAAGgtaagaaagaagatacagaagaatactttggtcatggcattcatgataaaattgtttcaaatat
 D  K  S

gaaaacatttacgtagcatttaatacgt<----3.3 kbp---->taattccagctactctggaggctgaggcaggaggatggcttgagccaggag

ttggaggctgcagttagccatgcttgtgccactacactcccagcccgggcaacagggcaagactctgtatctaaaaacaacaacaacaataata

gaaacaggtttcctttcccaagtttggaaaatctggtagtctctta
```

FIG. 11



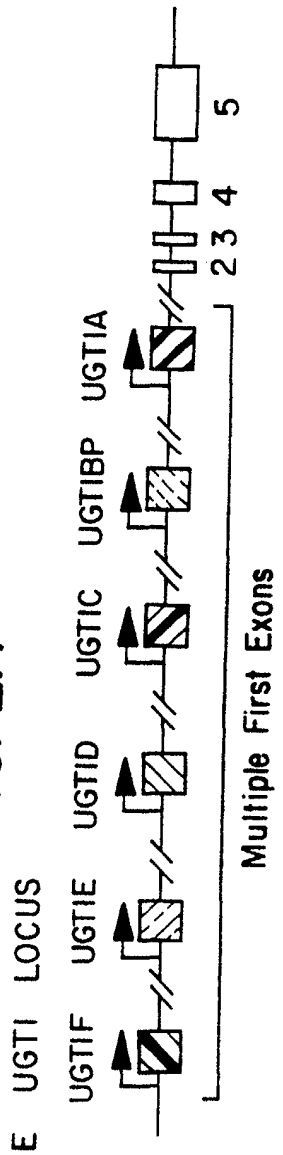

THE UGTI LOCUS
UGTIF
UGTIE
UGTID
UGTIC
UGTIBP
UGTIA
1
2 3 4
5
Multiple First Exons

FIG. 2A

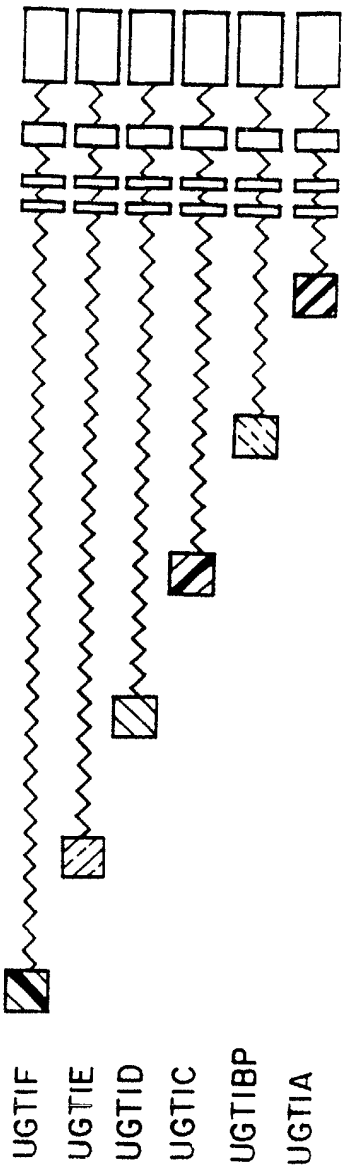

UGTIF
UGTIE
UGTID
UGTIC
UGTIBP
UGTIA

FIG. 2B

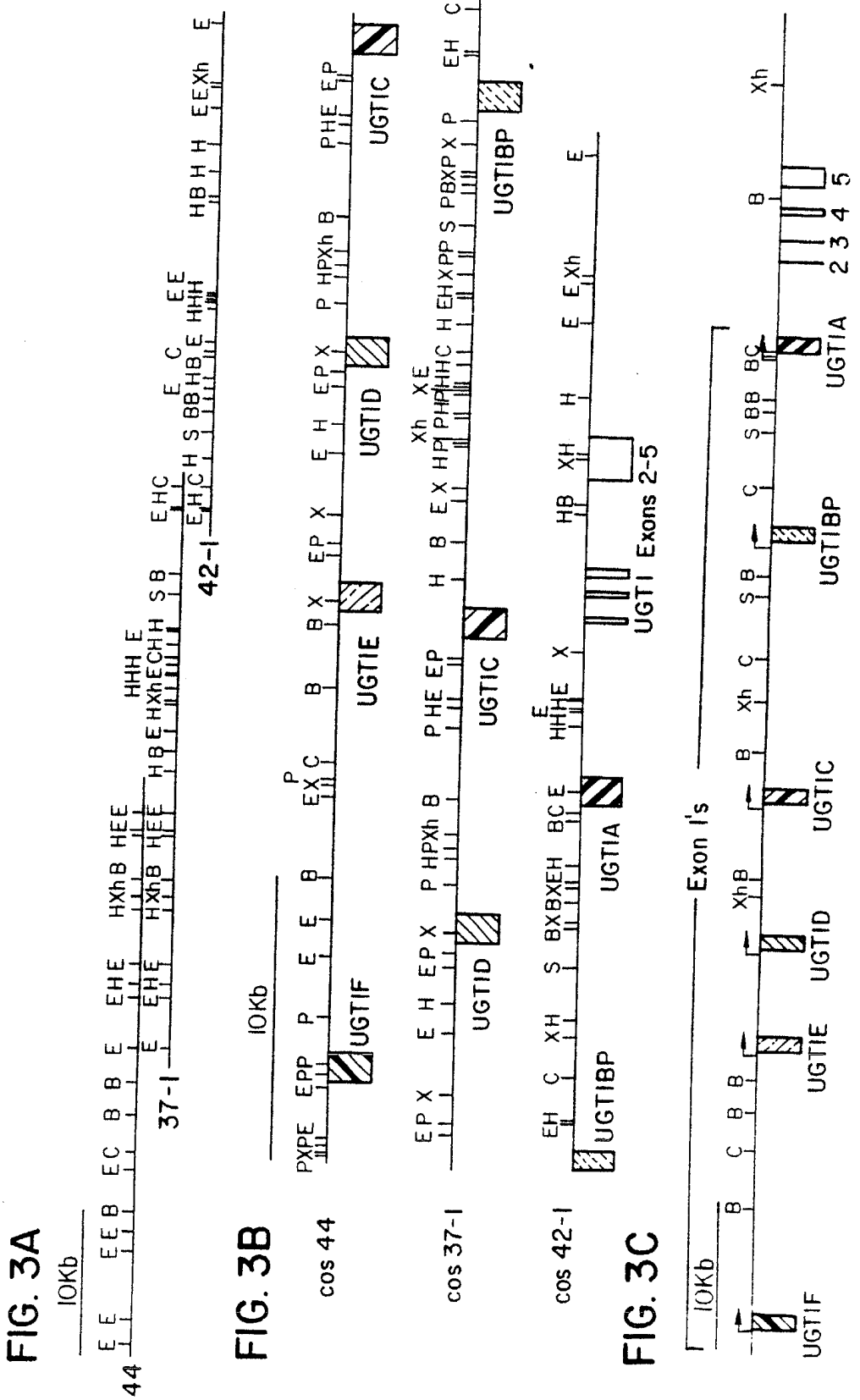


FIG. 3A
10Kb
44
37-1
42-1
FIG. 3B
10Kb
cos 44
UGTIF
UGTIE
UGTID
UGTIC
cos 37-1
UGTID
UGTIC
UGTIBP
cos 42-1
UGTIBP
UGTIA
UGT1 Exons 2-5
FIG. 3C
10Kb
Exon 1's
UGTIF
UGTIE
UGTID
UGTIC
UGTIBP
UGTIA
2 3 4 5

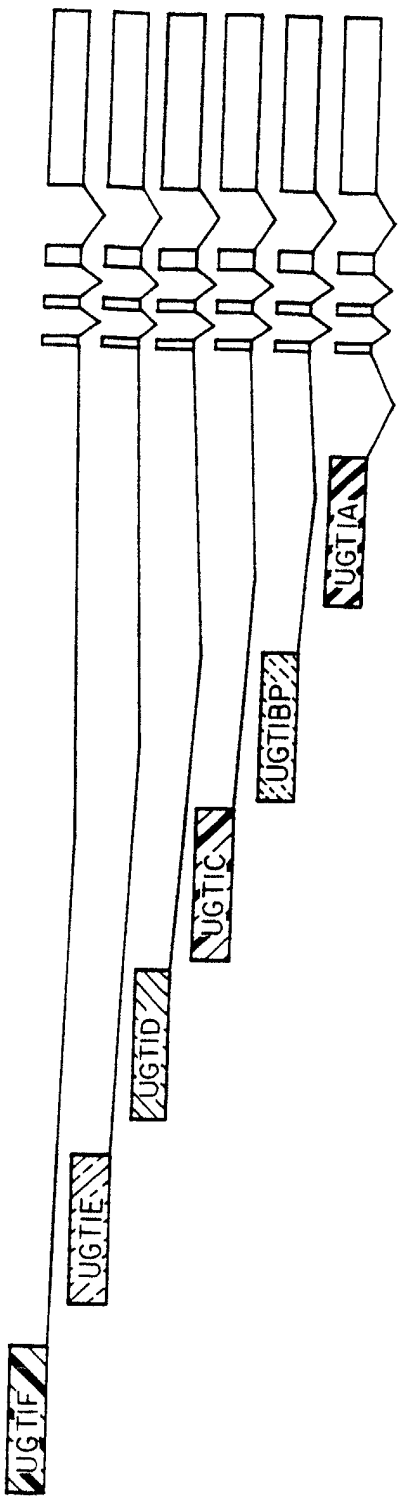


FIG. 3D

FIG. 3E

UGT1A GCCA**TATATATATATAATAA**GTAGGAGAGGGCCGAACCTCTGGCAGGAGCAAAGGCGCCC**ATG**

UGT1BP AGTGACAAGG**TAATAA**GATGAAGAAAACAAATGTAGGAGGC**CA**CAGCGTGGGGTGGACAGTCAGCTGTCCGTGTCTTCTGCTGAG**ATG**

UGT1C AGTGACAAGG**TAATAA**GATGAAGAAGCAAATGTAGCAGGC**CA**CAGCGTGGGGTGGACAGTCAGCTGTCCGTGTCTTCTGCTGAG**ATG**

UGT1D AATGACAAGG**TAATAA**GGCGAAGGAAACAAATGTAGCAGGC**CA**CAGCGTGGGGTGGACAGTCAGCTGTCGGTGGCTTCTGCTGAG**ATG**

UGT1E AGTGACAAGG**TAATAA**GACGAAGGAAACAATTCTAGGAGGC**CA**CAACGTGGGGTGGACAGTCAGCTGTCGGTGGCTTCTGCTGAG**ATG**

UGT1F GGTTCTTACA**TATCAAA**GGGTAAAATTCAGAGCAAGGGAGA**C**GTAGACAGGACCTGTGAAAAGCAGTGGTTAGTTT**A**TTTCCAGG**ATG**

AGGGAAAATACCTAGGAGCCCTGTGATTTGGAGAGTGAAAACTCTTTATTACCGTTGTTACTTTAACTC

FIG. 4A

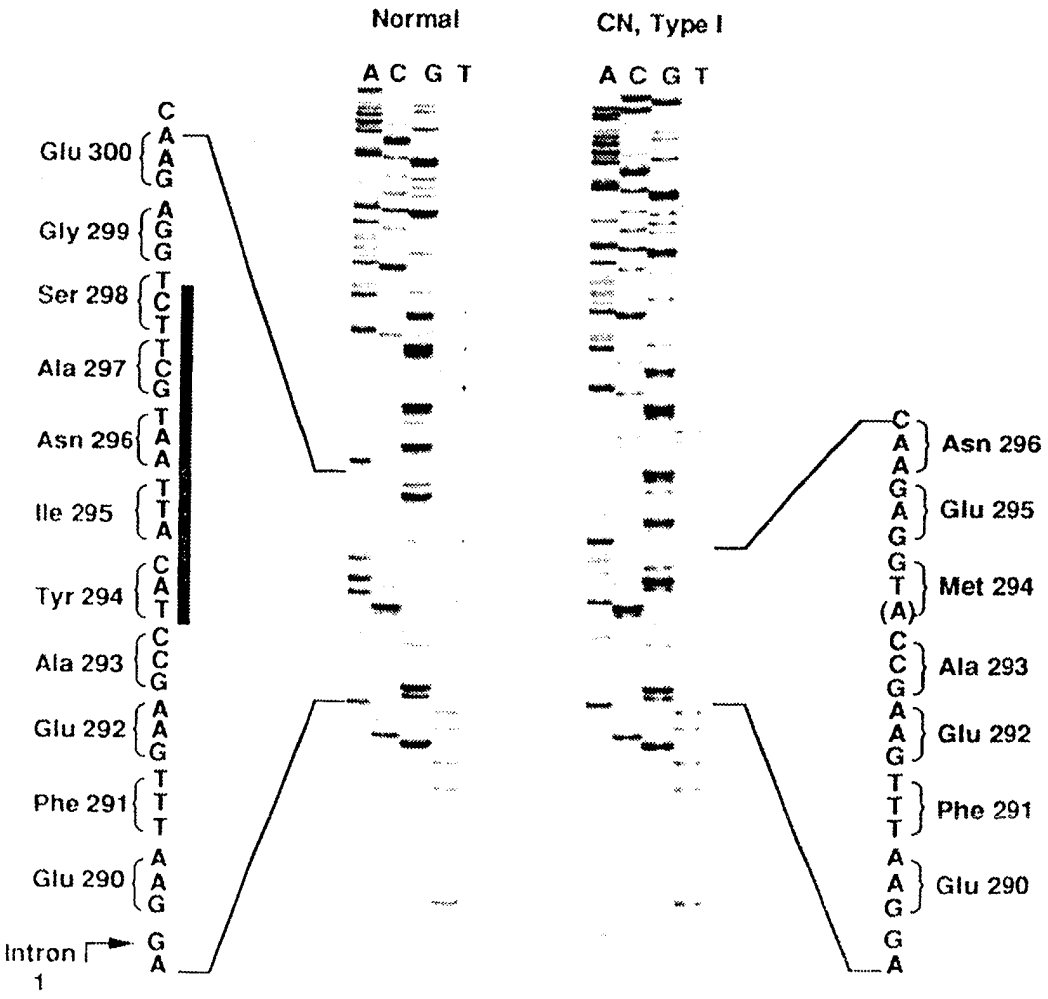

Normal
CN, Type I
A C G T
A C G T
Glu 300
Gly 299
Ser 298
Ala 297
Asn 296
Ile 295
Tyr 294
Ala 293
Glu 292
Phe 291
Glu 290
Intron 1
Asn 296
Glu 295
Met 294
Ala 293
Glu 292
Phe 291
Glu 290

FIG. 4B

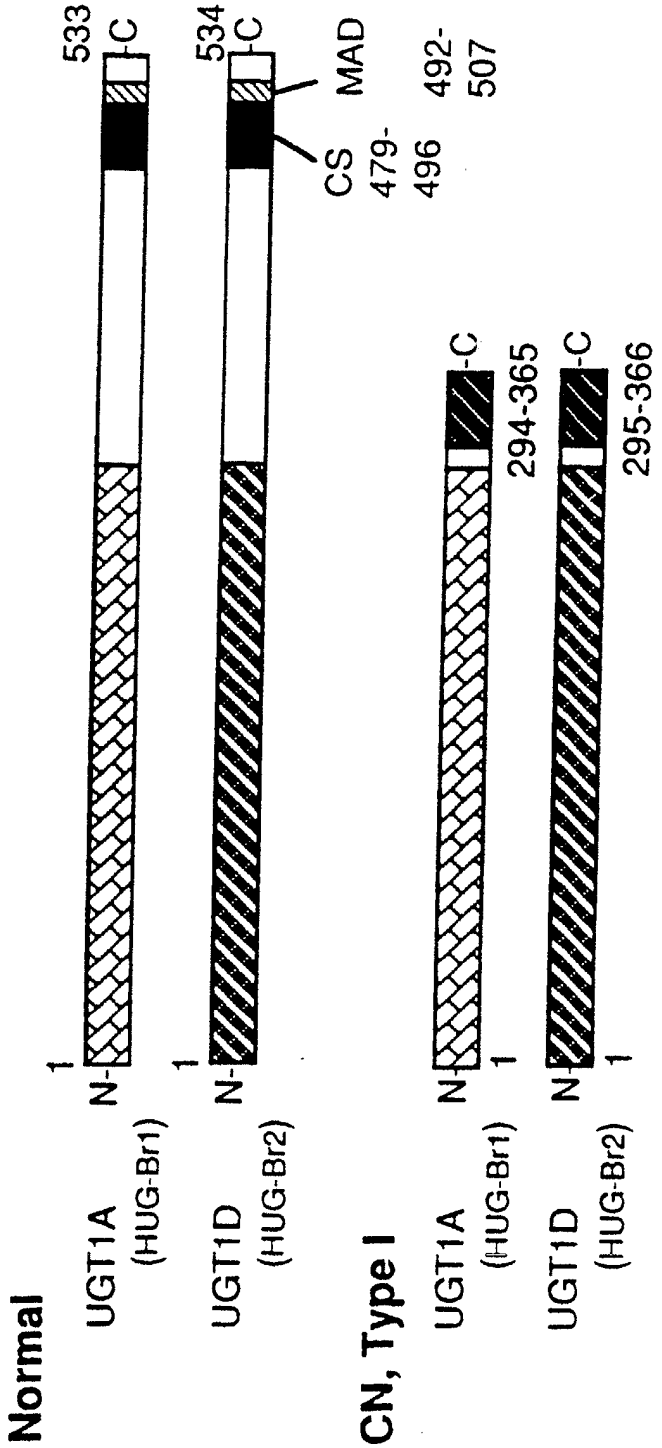
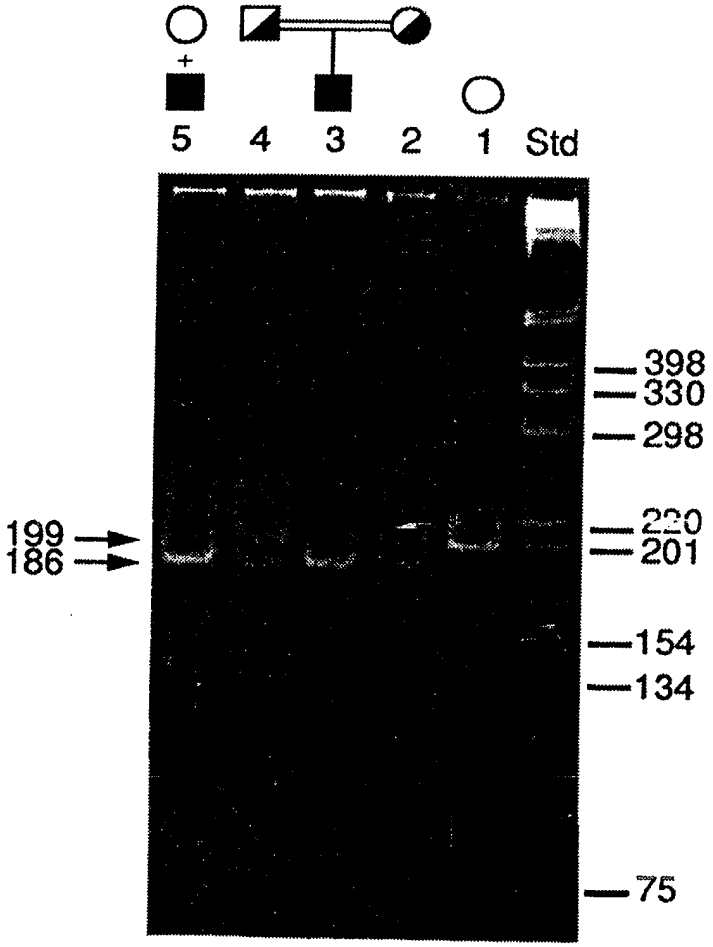
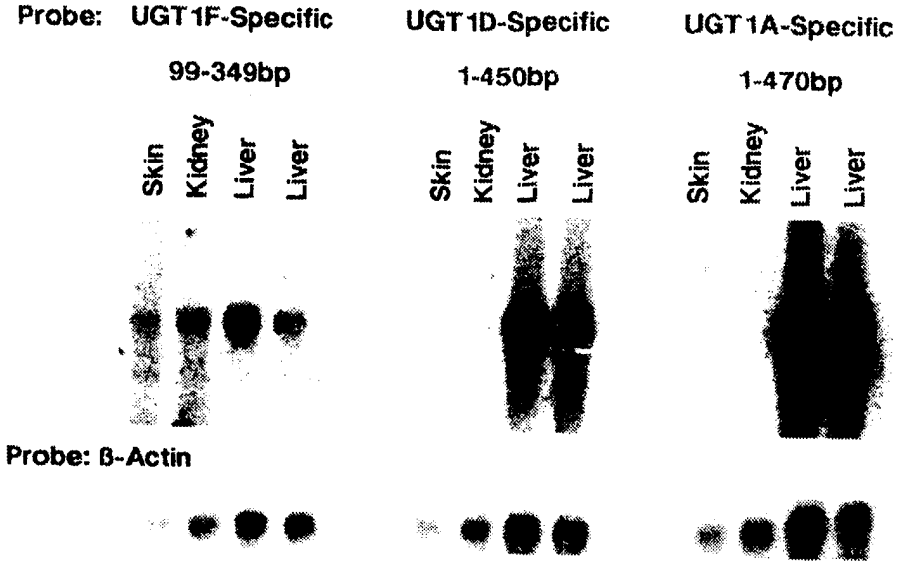

Normal
UGT1A
(HUG-Br1)
UGT1D
(HUG-Br2)
N-
C
1
533
534
CS
MAD
479-496
492-507
CN, Type I
294-365
295-366

FIG. 5



Northern

FIG. 6A



Southern

FIG. 6B

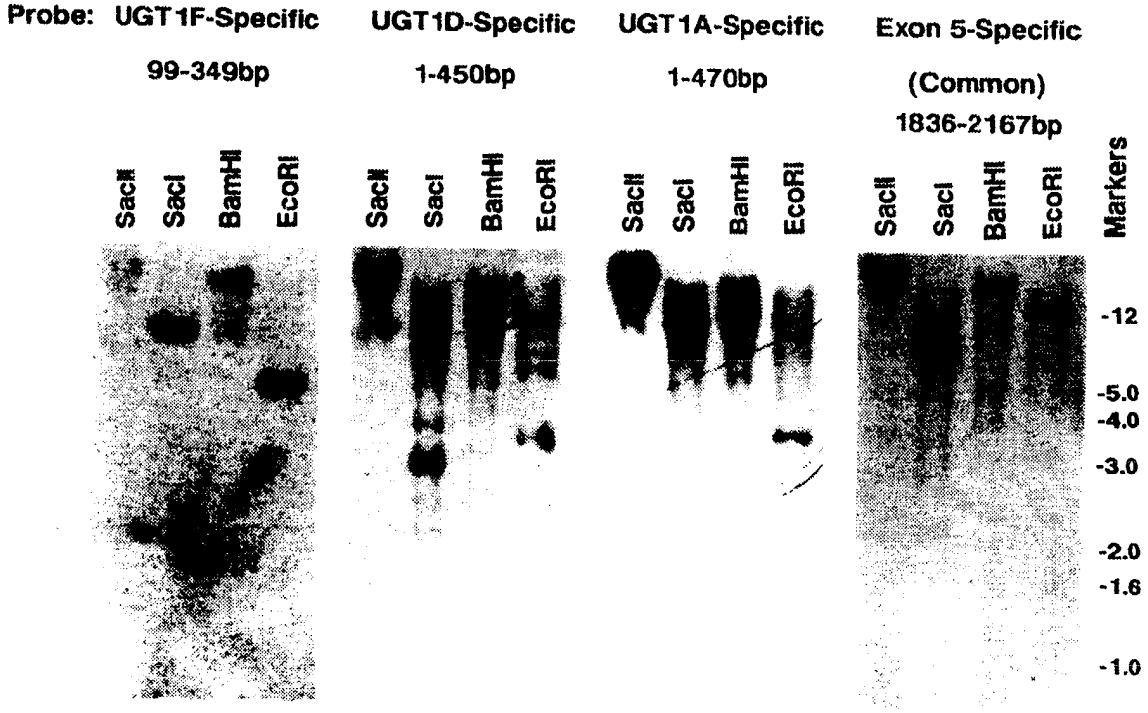


FIG. 7

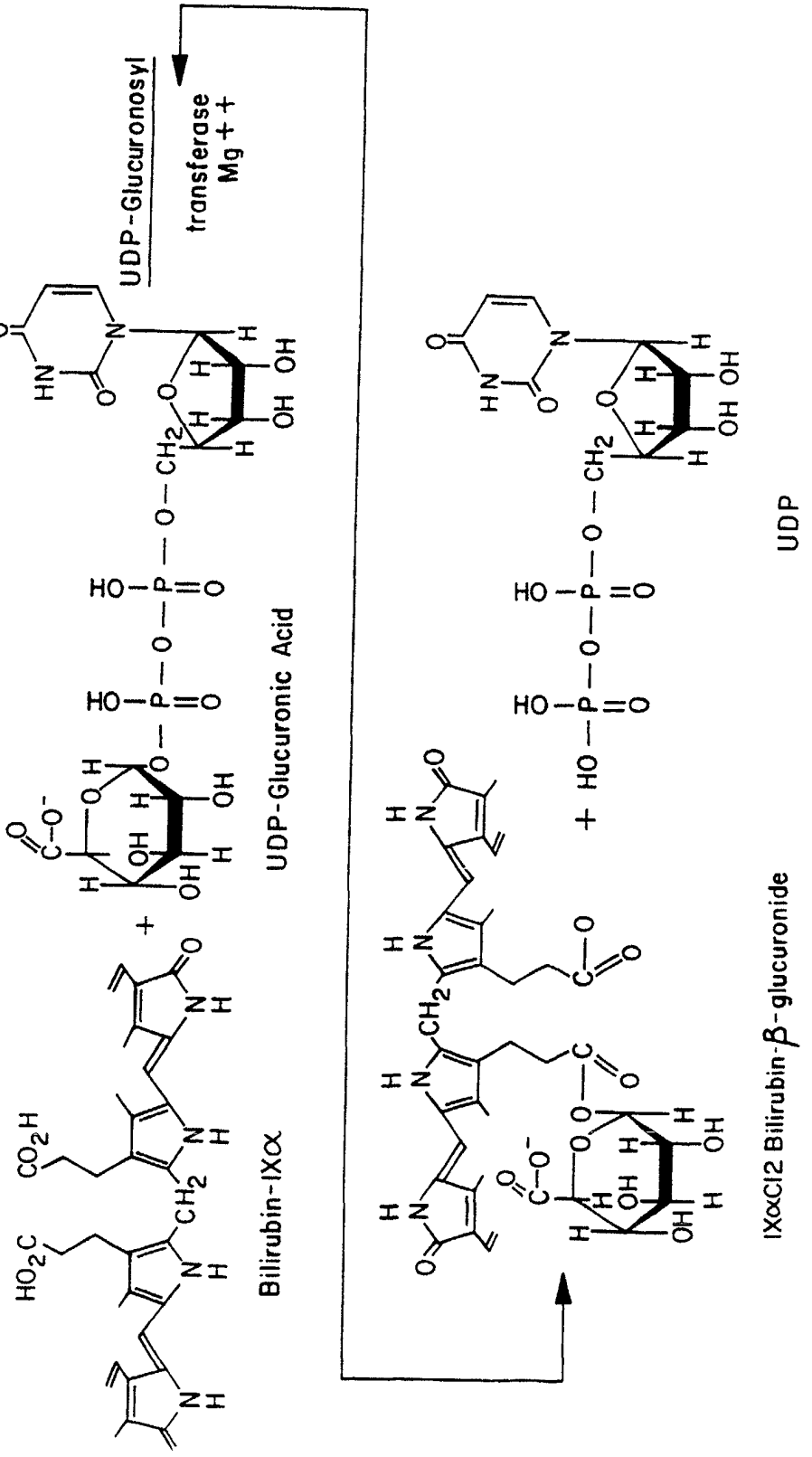
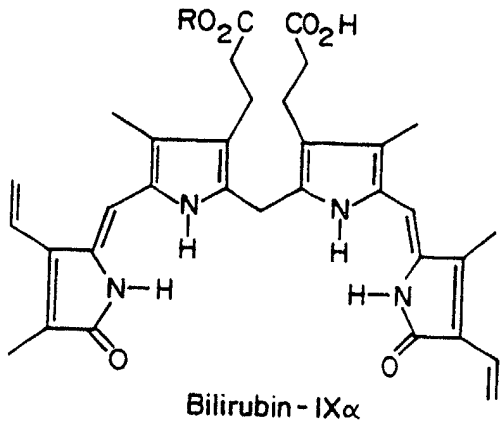
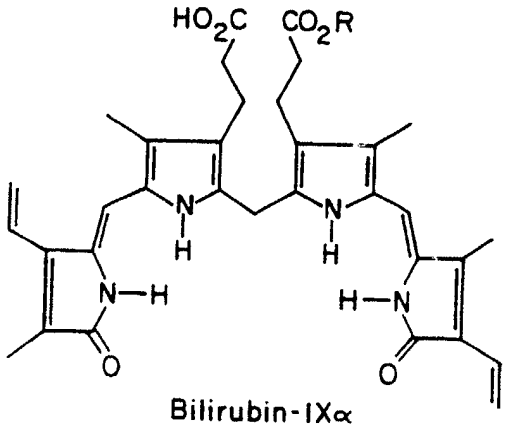

UDP-Glucuronosyl
transferase
Mg++
UDP-Glucuronic Acid
UDP
Bilirubin-IXα
IXαCl2 Bilirubin-β-glucuronide

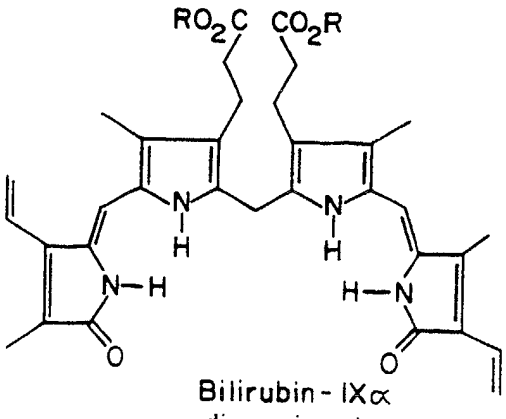
FIG. 8



Bilirubin-IXα
C-8 monoconjugat



Bilirubin-IXα
C-12 monoconjugate



Bilirubin-IXα
disconjugate

FIG. 9A

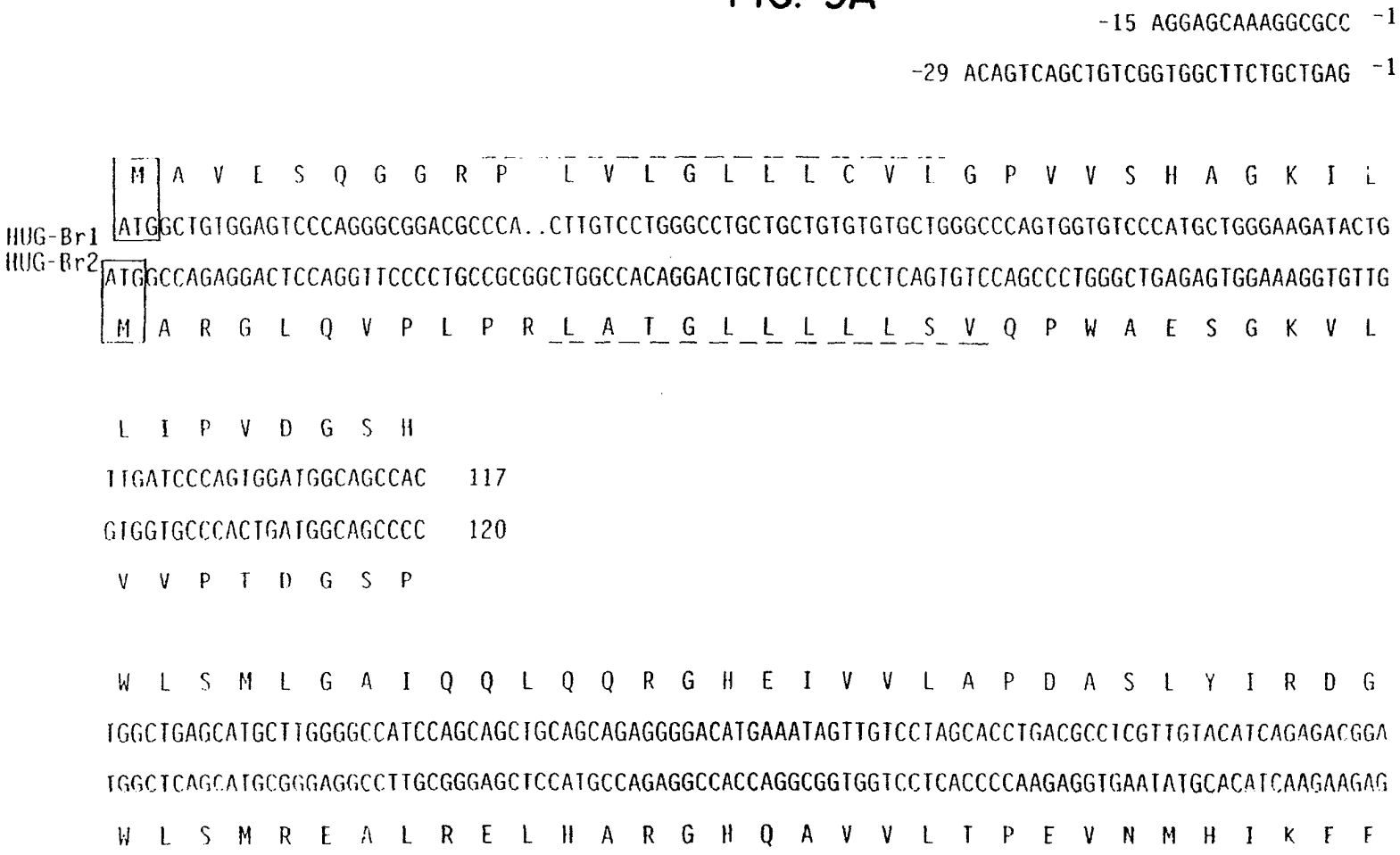


FIG. 9B

```
 A  F  Y  T  L  K  T  Y
GCATTTTACACCTTGAAGACGTAC   237
AAATTTTTCACCCTGACAGCCTAT   240
 K  F  F  T  L  T  A  Y

 P  V  P  F  Q  R  E  D  V  K  E  S  F  V  S  L  G  H  N  V  F  E  N  D  S  F  L  Q  R  V  I  K
CCTGTGCCATTCCAAAGGGAGGATGTGAAAGAGTCTTTTGTTAGTCTCGGGCATAATGTTTTTGAGAATGATTCTTTCCTGCAGCGTGTGATCAAAA
GCTGTTCCATGGACCCAGAAGGAATTTGATCGCGTTACGCTGGGCTACACTCAAGGGTTCTTTGAAACAGAACATCTTCTGAAGAGATATTCTAGAA
 A  V  P  W  T  Q  K  E  F  D  R  V  T  L  G  Y  T  Q  G  F  F  E  T  E  H  L  L  K  R  Y  S  R

 T  Y  K  K  I  K  K  D
CATACAAGAAAATAAAAAAGGAC   357
GTATGGCAATTATGAACAATGTA   360
 S  M  A  I  M  N  N  V
```

FIG. 9C

S A M L L S G C S H L L H N K E L M A S L A E S S F D V M L T D

TCTGCTATGCTTTTGTCTGGCTGTTCCCACTTACTGCACAACAAGGAGCTCATGGCCTCCCTGGCAGAAAGCAGCTTTGATGTCATGCTGACGGAC

TCTTTGGCCCTTCATAGGTGTTGTGTGGAGCTACTGCATAATGAGGCCCTGATCAGGCACCTGAATGCTACTTCCTTTGATGTGGTTTTAACAGAC

S L A L H R C C V E L L H N E A L I R H L N A T S F D V V L T D

P F L P C S P I

CCTTTCCTTCCTTGCAGCCCCATC 477

CCCGTTAACCTCTGCGGGGCGGTG 480

P V N L C G A V

V A Q Y L S L P T V F F L H A L P C S L E F E A T Q C P N P F S

GTGGCCCAGTACCTGTCTCTGCCCACTGTATTCTTCTTGCATGCACTGCCATGCAGCCTGGAATTTGAGGCTACCCAGTGCCCCAACCCATTCTCC

CTGGCTAAGTACCTGTCGATTCCTGCTGTGTTTTTTTGGAGGTACATTCCATGTGACTTAGACTTTAAGGGCACACAGTGTCCAAATCCTTCCTCC

L A K Y L S I P A V F F W R Y I P C D L D F K G T Q C P N P S S

FIG. 9D

```
 Y  V  P  R  P  L  S  S
TACGTGCCCAGGCCTCTCTCCTCT   597
TATATTCCTAAGTTACTAACGACC   600
 Y  I  P  K  L  L  Y  Y

 H  S  D  H  M  T  F  L  Q  R  V  K  N  M  L  I  A  F  S  Q  N  F  L  C  D  V  V  Y  S  P  Y  A
CATTCAGATCACATGACCTTCCTGCAGCGGGTGAAGAACATGCTCATTGCCTTTTCACAGAACTTTCTGTGCGACGTGGTTTATTCCCCGTATGCA
AATTCAGACCACATGACATTCCTGCAAAGGGTCAAGAACATGCTCTACCCTCTGGCCCTGTCCTACATTTGCCATACTTTTTCTGCCCCTTATGCA
 N  S  D  H  M  T  F  L  Q  R  V  K  N  M  L  Y  P  L  A  L  S  Y  I  C  H  T  F  S  A  P  Y  A

 T  L  A  S  E  F  L  Q
ACCCTTGCCTCAGAATTCCTTCAG   717
AGTCTTGCCTCTGAGCTTTTTCAG   720
 S  L  A  S  E  L  F  Q

 R  E  V  T  V  Q  D  L  L  S  S  A  S  V  W  L  F  R  S  D  F  V  K  D  Y  P  R  P  I  M  P  N
AGAGAGGTGACTGTCCAGGACCTATTGAGCTCTGCATCTGTCTGGCTGTTTAGAAGTGACTTTGTGAAGGATTACCCTAGGCCCATCATGCCCAAT
```

SUBSTITUTE SHEET

FIG. 9E

AGAGAGGTGTCAGTGGTGGATCTTGTCAGATATGCATCCGTGTGGCTGTTCCGAGGGGACTTTGTGATGGACTACCCAGGCCCGATCATGCCCAAC

R E V S V V D L V S Y A S V W L F R G D F V M D Y P R P I M P N

M V F V G G I N

ATGGTTTTTGTTGGTGGAATCAAC 837

ATGGTCTTCATTGGGGGCATCAAC 840

M V F I G G I N

FIG. 9F

C L H Q N P L S Q E F E A Y I N A S G E H G I V V F S L G S M V

TGCCTTCACCAAAATCCACTATCCCAGGAATTTGAAGCCTACATTAATGCTTCTGGAGAACATGGAATTGTGGTTTTCTCTTTGGAATCAATGGTC

TGTTGCCAACGGGAAGCCACTATCT

C A N G K P L S

S E I P E K K A 960

TCAGAAATTCCAGAGAAGAAAGCT

M A I A D A L G K I P Q T V L W R Y T G T R P S N L A N N T I L

ATGGCAATTGCTGATGCTTTGGGCAAAATCCCTCAGACAGTCCTGTGGCGGTACACTGGAACCCGACCATCGAATCTTGCGAACAACACGATACTT

V K W L P Q N D

GTTAAGTGGCTACCCCAAAACGAT 1080

L L G H P M T R A F I T H A G S H G V Y E S I C N G V P M V M M

CTGCTTGGTCACCCGATGACCCGTGCCTTTATCACCCATGCTGGTTCCCATGGTGTTTATGAAAGCATATGCAATGGCGTTCCCATGGTGATGATG

FIG. 9G

P L F G D Q M D

CCCTTGTTTGGTGATCAGATGGAC 1200

N A K R M E T K G A G V T L N V L E M T S E D L E N A L K A V I

AATGCAAAGCGCATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTGGAAATGACTTCTGAAGATTTAGAAAATGCTCAAAAAGCAGTCATC

N D K S Y K E N

AATGACAAAAGTTACAAGGAGAAC 1320

I M R L S S L H K D R P V E P L D L A V F W V E F V M R H K G A

ATCATGCGCCTCTCCAGCCTTCACAAGGACCGCCCGGTGGAGCCGCTGGACCTGGCCGTGTTCTGGGTGGAGTTTGTGATGAGGCACAAGGGCGCG

P H L R P A A H

CCACACCTGCGCCCCGCAGCCCAC 1440

FIG. 9H

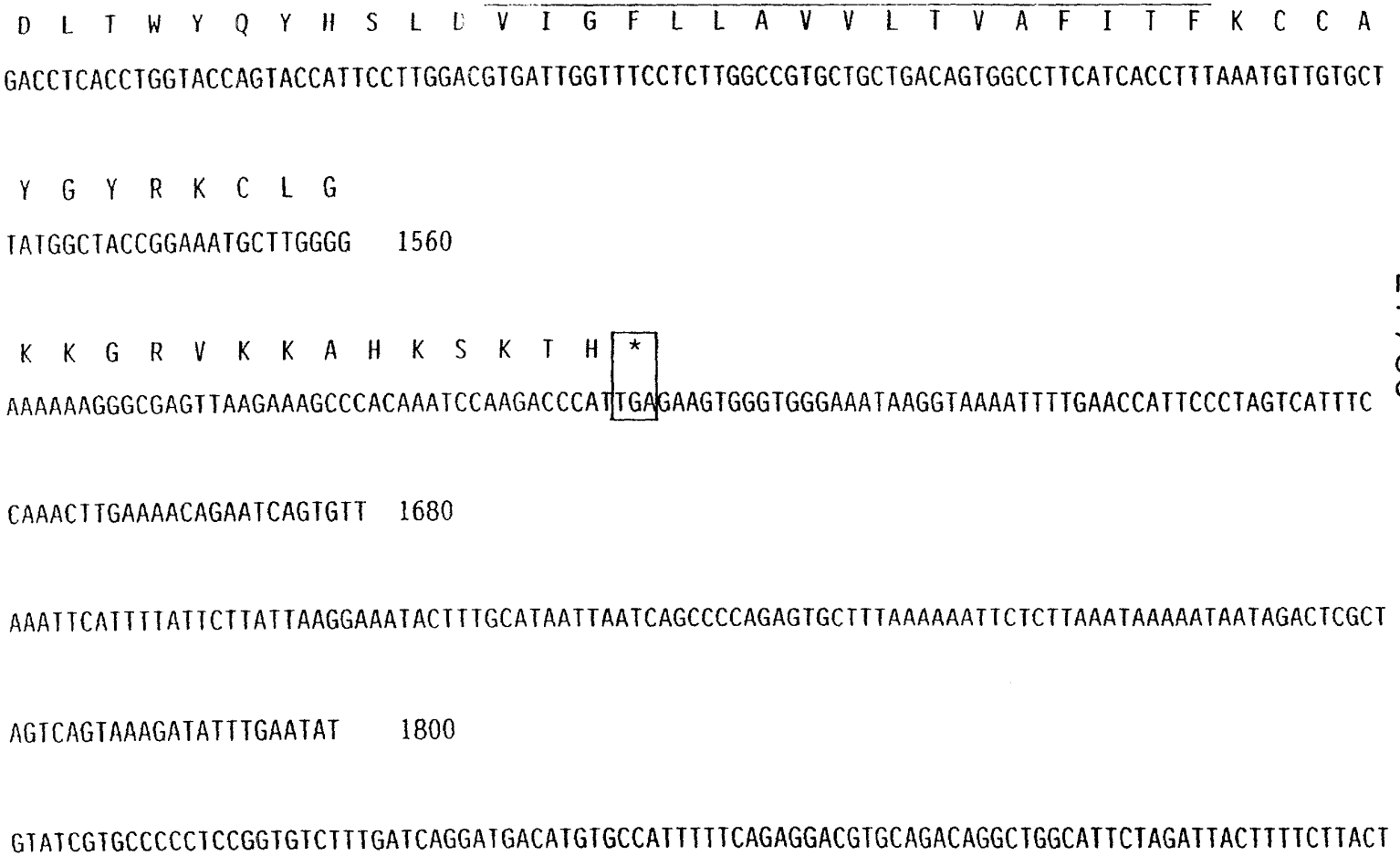


FIG. 9I

CTGAAACATGGCCTGTTTGGGAGT 1920

GCGGGATTCAAAGGTGGTCCCACCGCTGCCCCTACTGCAAATGGCAGTTTTAATCTTATCTTTTGGCTTCTGCAGATGGTTGCAATTGATCCTTAA

CCAATAATGGTCAGTCCTCATCTC 2040

TGTCCTGCTTCATAGGTGCCACCTTTGTGTGTTTAAAGAAGGGAAGCTTTGTACCTTTAGAGTGTAGGTGAAATGAATGAATGGCTTGGAGTGCAC

TGAGAACAGCATATGATTTCTTGCT 2160

TTGGGGAAAAAGAATGATGCTATGAAATTGGTGGGTGGTGTATTTGAGAAGATAATCATTGCTTATGTCAAATGGAGCTGAATTTGATAAAAACCC

AAAATACAGCTATGAAGTGCTGGG 2280

CAAGTTTACTTTTTTTCTGATGTTTCCTACAACTAAAATAAATTTAATAAATTTATAAA 2339

FIG. IOA

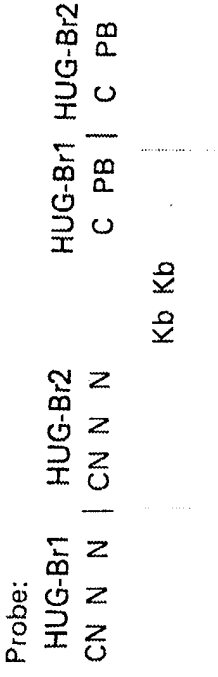


FIG. IOB

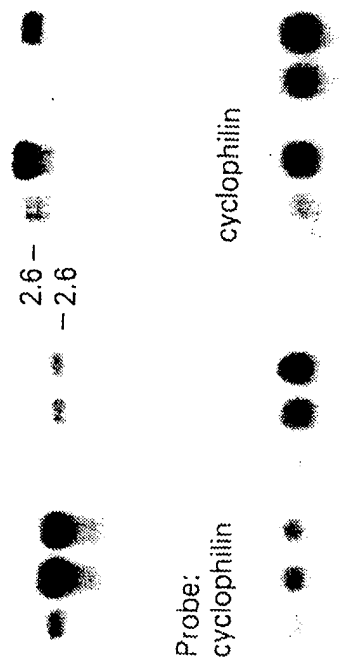


FIG. IIB

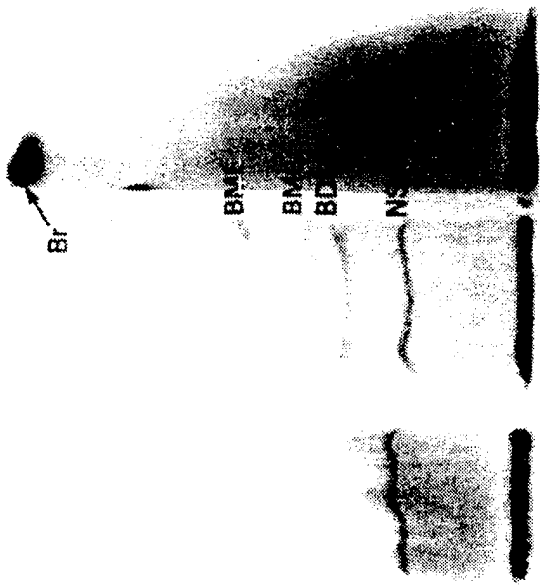


FIG. IIA

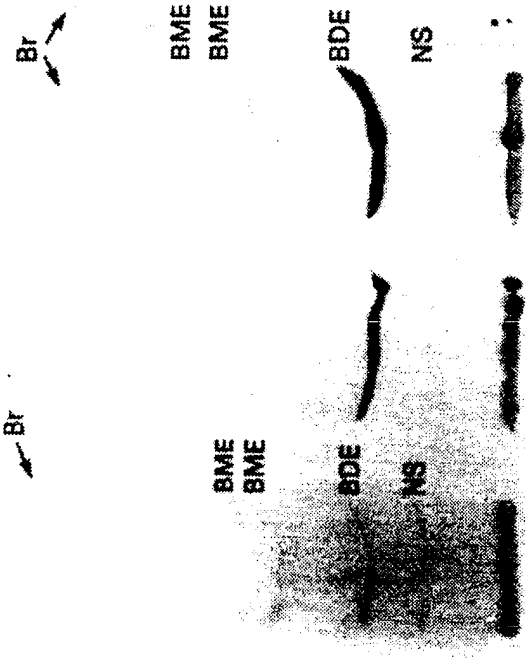


FIG. 12A

```
A Rat PHENOL MACLLPA.ARLPA.GFLFLVLWGSVL.GDKLLVVPQDGSHWLSMKEIVEHLSERGHDIVVLVPEVNLLLGESKYYRRKSFPVPYN
      LEELRTRYRSFGNNH  97

B     HLUGP1 MACLLRSFQRISA.GVFFLALWGMVV.GDKLLVVPQDGSHWLSMKDIVEVLSDRGHEIVVVVPEVNLLLKEYKYYTRKIYPVPYD
             QEELKNRYQSFGNNH  98

C     HUG-BR1 MAVESQGGRPLVL.GLLLCVLGPVVSHAGKILLIPVDGSHWLSMLGAIQQLQQRGHEIVVLAPDASLYIRDGAFYTLKTYPVPFQ
             REDVKESFVSLGHNV  99

D     HUG-BR2 MARGLQVPLPRLATGLLLLLSVQPWAESGKVLVVPTDGSPWLSMREALRELHARGHQAVVLTPEVNMHIKEEKFFTLTAYAVPWT
             QKEFDRVTLGYTQGF  100

E   Rat  Bil ..MGLCAPLRGLS.GLLLLLCALPWAEGGKVLVFPMEGSHWLSMRDVVRELHARGHQAVVLAPEVTVHMKGEDFFTLQTYAFPYT
             KEEYQREILGNAKKG  97


        A    FAASSPLMAPLREYRNNMIVIDMCFFSCQSLLKDSATLSFLRENQFDALFTDPAMPCGVILAEYLKLPSIYLFRGFPCSLEH.IG
             QSPSPVSYVPRFYTK  196

        B    FAERSFLTAPQTEYRNNMIVIGLYFINCQSLLQDRDTLNFFKESKFDALFTDPALPCGVILAEYLGLPSVYLFRGFPCSLEHTFS
             RSPDPVSYIPRCYTK  198

        C    FENDSFLQRVIKTYKKIKKDSAMLLSGCSHLLHNKELMASLAESSFDVMKTDPFLPCSPIVAQYLSLPTVFFLHALPCSLEFEAT
             QCPNPFSYVPRPLSS  199

        D    FETEHLLKRYSRSMAIMNNVSLALHRCCVELLHNEALIRHLNATSFDVVLTDPVNLCGAVLAKYLSIPAVFFMRYIPCDLDFKGT
             QCPNPSSYIPKLLTT  200

        E    FEPQHFVKTFFETMASIKKFFDLYANSCAALLHNKTLIQQLNSSSFDVVLTDPVFPCGALLAKYLQIPAVFFLRSVPCGIDYEAT
             QCPKPSSYIPNLLTM  197
```

FIG. 12B

A FSDHMTFPQRLANFIANILENYLYHCLYSKYEILASDLLKRDVSLPA.LHQNSLWLLRYDFVFEYPRPVMPNMIFIGGTNCKKKG
NLSQEFEAYVNASGE 295

B FSDHMTFSQRVANFLVNLLEPYLFYCLFSKYEKLASAVLKRDVDITT.LSEVSVWLLRYDFVLEYPRPVMPNMVFIGGINCKKRK
DLSQEFEAYINASGE 297

C HSDHMTFLQRVKNMLIAFSQNFLCDVVYSPYATLASEFLQREVTVQDLLSSASVWLFRSDFVKDYPRPIMPNMVFVGGINCLHQN
PLSQEFEAYINASGE 299

D NSDHMTFLQRVKNMLYPLALSYICHTFSAPYASLASELFQREVSVVDLVSYASVWLFRGDFVMDYPRPIMPNMVFIGGINCANGK
PLSQEFEAYINASGE 300

E LSDHMTFLQRVKNMLYPLTLKYICHILSITPYESLASELLQREMSLVEVLSHASVWLFRGDFVFDYPRPIMPNMVFIGGINCVIK
PLSQEFEAYVNASGE 297

A HGIVVFSLGSMVSEIPEKKAMEIAEALGRIPQTLLWRYTGTRPSNLAKNTILVKWLPQNDLLGHPKARAFITHSGSHGIYEGICN
VPMVMMPLFGDQMD 395

B HGIVVFSLGSMVSEIPEKKAMAIADALGKNPQTVLWRYTGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICN
GVPMVMMPLFGDQMD 397

C HGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRYTGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICN
GVPMVMMPLFGDQMD 399

D HGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRYTGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICN
GVPMVMMPLFGDQMD 400

E HGIVVFSLGSMVSEIPEKKAMEIAEALGRIPQTLLWRYTGTRPSNLAKNTILVKWLPQNDLLGHPKARAFITHSGSHGIYEGICN
GVPMVMMPLFGDQMD 397

FIG. 12C

A NAKRMETRGAGVTLNVLEMTADDLENALKTVINNKSYKENIMRLSSLHKDRPIEPLDLAVFWVEYVMRHKGAPHLRPAAHDLTWYQY HSLDVIGFLLAIV 495

B NAKRMETKGAGVLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVEPLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYH SLDVIGFLLAVV 497

C NAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVEPLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQY HSLDVIGFLLAVV 499

D NAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVEPLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQY HSLDVIGFLLAVV 500

E NAKRMETRGAGVTLNVLEMTADDLENALKTVINNKSYKENIMRLSSLHKDRPIEPLDLAVFWVEYVMRHKGAPHLRPAAHDLTWYQY HSLDVIGFLLAIV 497

A LTVVFIVYKSCAYGCRKCFGGKGRVKKSHKSKTH* 530

B LTVAFITFKCCPYGYPKCLGKKGRVKKAHKSKTH* 531

C LTVAFITFKCCAYGYRKCLGKKGRVKKAHKSKTH* 533

D LTVAFITFKCCAYGYRKCLGKKGRVKKAHKSKTH* 534

E LTVVFIVYKSCAYGCRKCFGGKGRVKKSHKSKTH* 531

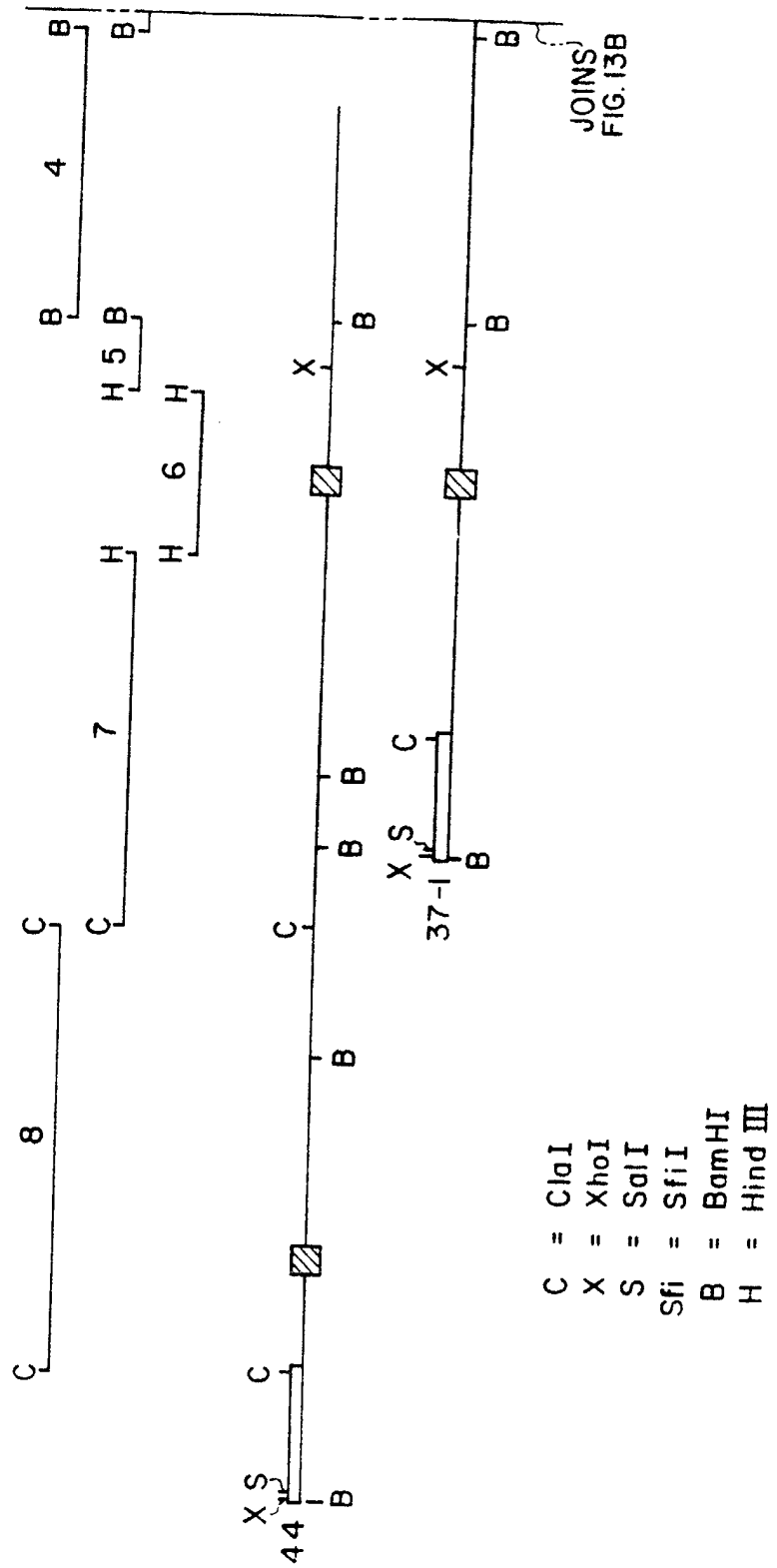

JOINS FIG. 13B
4
5
6
7
8
44
37-1
C = ClaI
X = XhoI
S = SalI
Sfi = SfiI
B = BamHI
H = HindIII

FIG. 13A

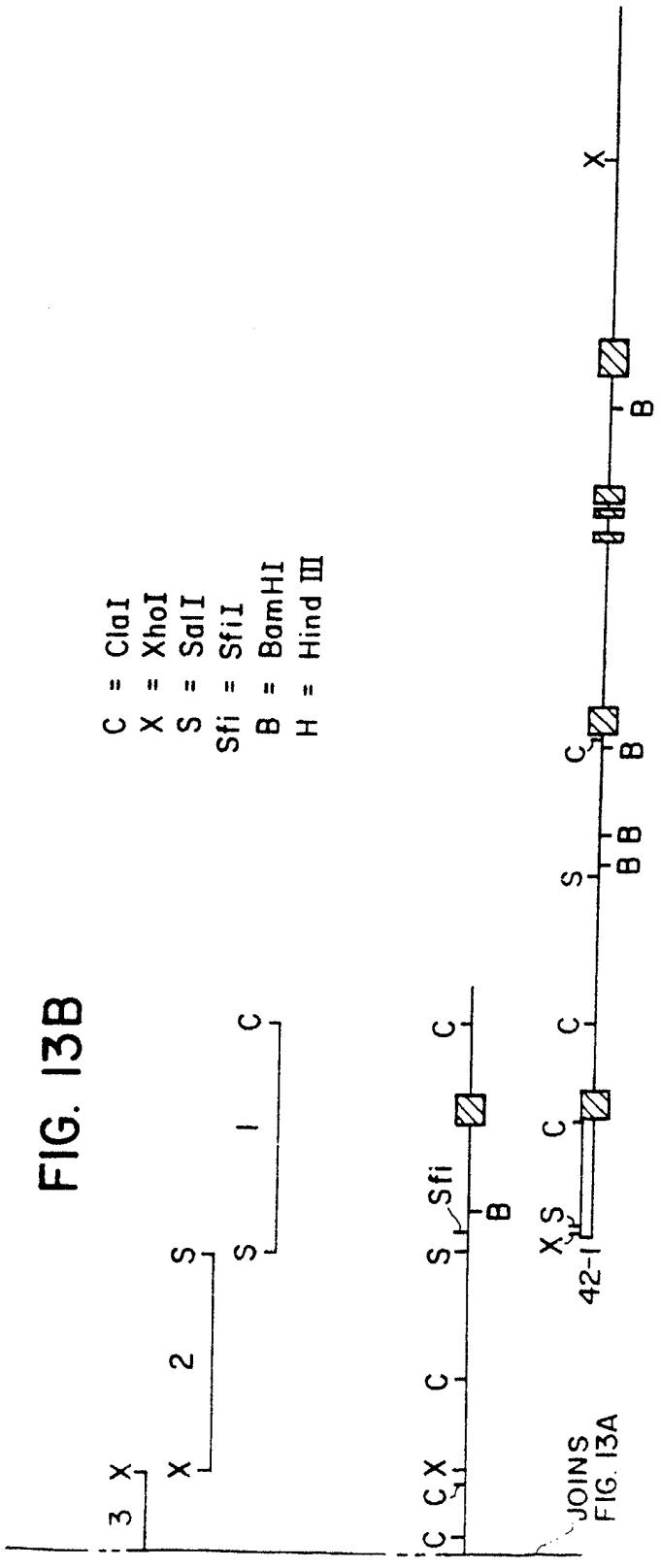

C = ClaI
X = XhoI
S = SalI
Sfi = SfiI
B = BamHI
H = Hind III
3
2
1
42-1
JOINS FIG. 13A

FIG. 13B

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US92/00282

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all)[3]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC (5): Please See Attached Sheet.
US CL : Please See Attached Sheet.

II. FIELDS SEARCHED

Minimum Documentation Searched[4]

Classification System	Classification Symbols
U.S.	536/26, 27, 28, 29; 435 6, 91, 240.2, 252.3, 320.1; 530/395; 436/501, 94; 935/22, 66, 77, 78

Documentation Searched other than Minimum Documentation to the extent that such Documents are included in the Fields Searched[6]

Please See Attached Sheet.

III. DOCUMENTS CONSIDERED TO BE RELEVANT[14]

Category*	Citation of Document,[16] with indication, where appropriate, of the relevant passages[17]	Relevant to Claim No.[18]
Y	Proceedings of the National Academy of Sciences, Vol. 85, No. 22, Issued 1988, Harding et al., "Cloning and Substrate Specificity of a Human Phenol UDP-Glucuronosyltransferase Expressed in COS-7 Cells - DNA Sequence Determination", pages 8381-8385, see entire document.	1-36
Y	Federation of European Biological Societies Letters, Vol. 243, No. 2, Issued 1989, Fournell-Gigleux et al., "Expression of a Human Liver cDNA Encoding a UDP-Glucuronosyltransferase Catalyzing the Glucuronidation of Hyodeoxycholic Acid in Cell Culture", pages 119-122, see entire document.	1-36
Y	Journal of Biological Chemistry, Vol. 265, No. 14, Issued 1990, Ritter et al., "Cloning and Expression of Human Liver UDP-Glucuronosyltransferase in COS-1 Cells, 3',4'-Catechol Estrogens and Estrol as Primary Substrates", see entire document.	1-36

* Special categories of cited documents:[15]

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search[2]	Date of Mailing of this International Search Report[2]
09 APRIL 1992	26 MAY 1992
International Searching Authority[1]	Signature of Authorized Officer[20]
ISA/US	Stephanie W. Zitomer, Ph.D.

FURTHER INFORMATION CONTINUED FROM THE SECOND SHEET

V. ☐ **OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE**[1]

This international search report has not been established in respect of certain claims under Article 17(2) (a) for the following reasons:

1. ☐ Claim numbers _, because they relate to subject matter (1) not required to be searched by this Authority, namely:

2. ☐ Claim numbers _, because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out (1), specifically:

3. ☐ Claim numbers _, because they are dependent claims not drafted in accordance with the second and third sentences of PCT Rule 6.4(a).

VI. ☒ **OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING**[2]

This International Searching Authority found multiple inventions in this international application as follows:

Please See Attached Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application.

2. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims:

3. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers:

4. ☒ As all searchable claims could be searched without effort justifying an additional fee, the International Search Authority did not invite payment of any additional fee.

Remark on protest

☐ The additional search fees were accompanied by applicant's protest.

☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PREVIOUS SHEETS

I. CLASSIFICATION OF SUBJECT MATTER:
IPC (5):

C07H 15/12; C07K 15/14; G01N 33/48, 33/566; C12Q 1/68; C12P 19/34; C12N, 5/00, 15/00, 1/20

I. CLASSIFICATION OF SUBJECT MATTER:
US CL :

536/26, 27, 28, 29; 435 6, 91, 240.2, 252.3, 320.1; 530/395; 436/501, 94; 935/22, 66, 77, 78

II. FIELDS SEARCHED
Other Documents Searched:

Dialog Onesearch: Biosis, Biotech Abstracts, Embase, Medline, WPI, Diss. Abstracts, CAS; APS; Search terms: UDP-glucuronosyltransferase, recomb?, cDNA, probe, hybridiz?; Sequence searched by STIC staff

VI. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
This ISA found multiple inventions as follows:

I. Claims 1, 2, 9, 11-14 and 29-34, drawn to a first product classified in Class 536, subclass 27 and to a first process of using said product, classified in Class 435 subclass 6;
II. Claims 3-8, 10 and 15-18, drawn to a second product, classified in Class 435, subclass 320.1;
III. Claims 19-23, drawn to a third product, classified in Class 435, subclass 252.3;
IV. Claims 24-28, drawn to a fourth product, classified in Class 435, subclass 240.1;
V. Claims 35 and 36, drawn to a fifth product, classified in Class 530, subclass 395.
Unity is lacking in view of the additional multiple, structurally distinct inventions in the instant case which are not provided for by PCT Rules 13.1-13.4.